

# **Statistical Analysis of the Benefits of Outdoor Adventure Activities Programmes**

**Roger Antony Scrutton**

**Thesis submitted for the degree of  
Master of Philosophy  
University of Edinburgh  
2011**

## **Declaration**

I, Roger Antony Scrutton, declare that this thesis has been composed solely by myself and that it has not been submitted in whole or in part for any other degree or professional qualification. Except where otherwise acknowledged, the research is entirely my own.

Signed: .....

Date: .....

## **ABSTRACT**

There is a long-standing call for robust empirical evidence that a residential experience of outdoor adventure activities (OAA) is beneficial for the personal and social development (PSD) of young people. This call reflects the fact that consistently strong anecdotal and qualitative evidence of positive benefit is not mirrored in quantitative measures. To date, statistically significant measures of benefit, almost entirely made outside Scotland, range from negative to medium positive effect sizes.

This quantitative research measures the benefit of a residential week of introductory OAA for the PSD of upper primary school children, with the overarching aim of contributing to the argument for full funding of a “residential” for every Scottish school pupil. In the absence of suitable questionnaires, a questionnaire tailored to the project was created and tested for validity and reliability. Despite poor reliability, but encouraged by the repeatability of the distribution of pupils’ mean scores across administrations, the research was continued. Scores from over 300 pupils yielded a small, statistically significant improvement in their perception of their PSD skills from immediately before to immediately after their “residential”, but then a large decline measured two-three months later.

Underlying this trend, low-scoring pupils perceived greater benefit and retained this into the delayed test, girls demonstrated a more positive view of their skills than boys, and all pupils had greater belief in their social competence than in their confidence. Socio-economic data clearly supported the presence of a school effect, there was some evidence of a euphoria effect around the time of the “residentials”, and some evidence of improved understanding of the concepts embodied in the questionnaire over successive tests.

Regarding the implications of these results for research methodology, it is suggested that inconsistency in quantitative measures of benefit relates, in part, to the presence of numerous variables and survey effects that are often poorly constrained. For teachers, it is suggested that the observed loss of benefit 2-3 months later might be overcome if the experience is better integrated into the curriculum. For education policy makers and funders, the implication is that only if the experience is available free of charge to all pupils will it be fully integrated and its full benefit realised. Outdoor centre managers are encouraged to maintain quantitative records of pupils’ and teachers’ feedback on benefit to strengthen the case for full funding.

## EXTENDED ABSTRACT

There is a long-standing call for robust empirical evidence that a residential experience of outdoor adventure activities (OAA) is beneficial for the personal and social development (PSD) of young people. Strong anecdotal and qualitative evidence of positive benefit has been sufficient to see such experiences offered by schools for some time, but on an optional basis and heavily subsidised by parents. Now, the outdoor learning component of the Curriculum for Excellence offers a more formal umbrella under which residential weeks might be included in the curriculum. It is estimated that at a cost of £10M per year every school pupil in Scotland could enjoy a residential week free of charge. However, more quantitative measures of positive benefit are needed in order to justify improved funding of this sort of course in order for it to be included in the curriculum.

The search for rigorous and robust empirical evidence of benefit for PSD dates from the 1960s and by the 1990s had created sufficient volume to warrant meta-analysis. Measured benefit for the PSD of young people ranged from negative effects to small or medium, statistically significant positive effects. However, few studies targeted late 10-12 year old children, who are those offered the option in Scottish primary schools. In a recent study of a geography field course by Nundy (1999), that did target this age group, benefit was measured in pupils' perception of their academic skills but not their personal and social skills. In fact, at this time in research, we see a worrying distinction between wholly positive anecdotal and qualitative evidence for the positive benefits of OAA but only mixed results from quantitative research.

Against this background, the quantitative research reported here specifically considers the benefits of a residential week of introductory OAA for the PSD of upper primary school children. This is with the overarching aim of contributing to the argument for full funding of a residential visit for every child. In the absence of suitable published questionnaires, a questionnaire tailored to the project was created and tested for validity and reliability. Although results from the questionnaire yielded a consistent picture from application to application, formal measures of reliability indicated that it was barely reliable, with Spearman's correlation coefficients typically 0.4 and Cronbach alpha values of up to 0.8. (Note that the data were negative skewed). Despite this poor result, but encouraged by the repeatability of the distribution of pupils' mean scores, the research was pursued. A Control Experiment with no intervention was conducted with which the results of the Main Study

could be compared, but the schedule of the project required this to be done with pupils (117) from a different group of schools attending a different outdoor centre. The Main Study was conducted with 368 pupils from nine schools that attended Lagganlia, Dounans and Broomlee Outdoor Centres. It yielded a small but statistically significant improvement in the pupils' perception of their PSD skills from immediately before to immediately after their residential experience, not seen in the Control Experiment, but then a large decline in perception measured in a delayed test two-three months later. This overall trend hides the fact that low-scoring pupils perceived greater personal benefit at the time of the adventure week and retained this into the delayed test. Girls demonstrated a more positive view of their skills than boys in all aspects of the tests in the Main Study, but not in the Control Experiment. For both boys and girls, scores were higher in the Main Study than in the Control Experiment. Socio-economic data support this being a school group effect; alternatively, it could have been a euphoria effect on the pupils visiting the outdoor centres. If the latter, the reduced scores in the delayed test, which were approximately the same as the Control Experiment scores, might reflect a return to a more realistic base level of self perception amongst pupils in the Main Study. The fall in scores in the delayed test could also be attributed to class teachers doing little or no follow-up work on the residential experience.

The questionnaire contained five groups of questions (effectively, intended factors), which permitted the analysis by regression of how pupils' scores in the Main Study following the intervention depended on scores before. One factor was "Awareness of the Outdoor Environment" and the other four were on aspects of PSD, two personal – "Perception of Competence" and "Perception of Confidence" – and two social – "Awareness of Others" and "Attitude towards Groups". Although a factor analysis of the questionnaire did not readily identify these groups of questions, they did yield the same pattern of scores from pupils across all administrations of the questionnaire. It would be wise to consider the following outcomes with some caution. Linear multivariate regression modelling of the dependence of self-perceived PSD skills following the residential week of OAA clearly demonstrated that a pupil's Awareness of the Outdoor Environment did not influence their ability to benefit in PSD areas. The modelling showed that, for boys, the strongest predictor of benefit in PSD was Attitude towards Groups, whereas for girls the most important predictor was Perception of Competence. This pattern changed when scores immediately after the intervention were used to predict the delayed test scores, but across all tests, Perception of Competence

emerged as a reasonably strong predictor while Perception of Confidence was a relatively weak predictor.

In considering the implications of these results for research methodology, it is suggested that the rather mixed nature of quantitative measures of benefit relates, in part, to the presence in any project of many variables that are often poorly constrained. Making every effort to constrain these variables and avoid the presence of unwanted effects, such as a school group effect or a euphoria effect, is essential. For teachers, the principal implication is that maximum benefit will be obtained if the experience is integrated into the curriculum. Unfortunately, teachers are constrained by the fact that the experience is not an option for their pupils who come from families who cannot afford to sponsor their child. Therefore, for policy makers and funders, the message is that only if the experience is available free of charge to all pupils will its full impact be realised. Outdoor centre managers are encouraged to maintain quantitative records of pupils' and teachers' feedback on benefit to strengthen the case for full funding.

# CONTENTS

	<b>Page</b>
<b>Abstract</b>	<b>i</b>
<b>Extended Abstract</b>	<b>ii</b>
<b>List of Figures</b>	<b>viii</b>
<b>List of Tables</b>	<b>x</b>
<b>Acknowledgements</b>	<b>xii</b>
<b>Chapter 1     Introduction</b>	<b>1</b>
1.1     Background to this project	1
1.2     Outdoor Adventure Activities and Personal and Social Development	3
1.3     The search for a quantitative measure	4
1.3.1     Effect size	5
1.4     The nature of this study	9
1.5     Structure of this thesis	11
<b>Chapter 2     Experimental and questionnaire design</b>	<b>12</b>
2.1     Introduction	12
2.2     Project structure	13
2.2.1     Variables, constants and assumptions	14
2.2.2     Project schedule, the sample and data collection	15
2.3     Questionnaire development	17
2.3.1     Rationale	18
2.3.2     Questionnaire validity and reliability	19
Initial validity	20
Pilot Study reliability and its implications	22
Further questionnaire development	26
Questionnaire reliability in the Main Study	30
2.4     Summary of chapter	33
<b>Chapter 3     Data analysis</b>	<b>35</b>
3.1     Introduction	35
3.2     Data management and software	36
3.3     Data inspection and descriptives	36
3.3.1     Distribution and significance of scores on items	37

	Distribution of scores on items	37
	Statistical significance of scores on items	42
3.3.2	Distribution and significance of scores by pupils	44
	Pupils' scores on Personal and Social Development items	44
	Statistical significance of PSD scores by pupils	50
	Pupils' scores on "Awareness of the Outdoor Environment" items	53
3.4	Regression analysis	57
3.4.1	Univariate regressions	57
	Univariate regressions between Pre-test, Post-test and Delayed-test scores	57
	Factors influencing regression results	59
	Avoiding the factors influencing regression results	62
	Univariate regressions using individual PSD factors	65
	Summary of Section 3.4.1	66
3.4.2	Multivariate regression	67
	Initial regression analysis	68
	Next-step regression analysis	72
	Hierarchical regression analysis	75
	Summary of Section 3.4.2	76
3.4.3	Extension of regression analysis to the Main Study Delayed-test	76
	Initial regression analysis	77
	Next-step regression analysis	80
	Hierarchical regression analysis	83
	Summary of Section 3.4.3	84
3.5	Summary of chapter	85
<b>Chapter 4</b>	<b>Implications for research methodology</b>	<b>89</b>
4.1	Introduction	89
4.2	Conflicting research outcomes	89
4.2.1	The importance of variables	90
4.2.2	Consequences of the measurement of variables	91
4.2.3	Defining variables in the questionnaire	94
4.2.4	Capping effects on a Likert scale	96
4.3	Interesting but unwanted effects	98
4.3.1	A school group effect	98
4.3.2	A euphoria effect?	99
4.3.3	The importance of children's learning processes	103



<b>Chapter 5</b>	<b>Implications for practitioners</b>	<b>104</b>
5.1	Introduction	104
5.2	Implications for teachers	104
5.3	Implications for outdoor centres	107
5.4	Implications for policy makers	112
<b>Chapter 6</b>	<b>Summary of project outcomes</b>	<b>114</b>
6.1	Summary of key research results	114
6.1.1	Summary of results from Chapter 2	115
6.1.2	Summary of results from Chapter 3	116
6.2	Summary of implications	120
6.2.1	Summary from Chapter 4	120
6.2.2	Summary from Chapter 5	121
<b>References</b>		<b>123</b>
<b>Appendices</b>		<b>126</b>
Appendix 2.1	The questionnaire	126
Appendix 3.1	Histograms of scores on items	130
Appendix 3.2	Histograms of the distribution of pupils' mean scores	161
Appendix 3.3	Effect sizes and statistical significance of pupils' mean scores in factors	178
Appendix 3.4	Linear regression graphs for Pre-test vs Post-test vs Delayed-test mean scores in the Main Study and Control Experiment, for all cases and for boys and girls	182
Appendix 4.1	An understanding of children's learning process during outdoor adventure activities	195

## List of figures

Figure 1.1	Distribution of all effect sizes extracted by Hattie et al. (1997) from 96 studies. These are for all age groups on all programmes. Mean = 0.34. N.B. many negative effects	Page 6
Figure 2.1	Left: strong linear relationship between mean scores on the 30 questionnaire items, Pilot Study. Right: similar graph for the mean scores of 100 pupils	Page 22
Figure 2.2	Left: strong linear relationship between mean scores on the 30 questionnaire items, Control Experiment. Right: similar graph for the mean scores of 117 pupils	Page 28
Figure 3.1	Reference distributions of scores on items. Left: the count of scores increases progressively from 1 to 4 on the Likert scale; Right: 3 is the modal score	Page 38
Figure 3.2	Column charts of changes in mean scores on items across the five applications of the questionnaire: Pre-, Post- and Delayed-test of the Main Study and Pre- and Post-test of the Control Experiment, grouped by intended factor. Note that the two experiments were conducted with pupils from two different groups of schools	Page 40
Figure 3.3	Distribution of pupils' mean scores on PSD items for the Main Study and Control Experiment, for all cases and for boys and girls. SD = standard deviation; N = size of sample. N.B. the Count (vertical) scale changes with sample size	Page 45
Figure 3.4	Changes in pupils' mean PSD scores across tests	Page 53
Figure 3.5	Changes in pupils' mean scores on Awareness of the Outdoor Environment items across tests	Page 55
Figure 3.6	Univariate linear regressions for the dependence of change in pupils' mean scores between tests in the Main Study, using all cases and all items: a) data used for the regressions in this figure, copied from Appendix 3.4; b) SPSS scatter plot of dependence of change in score Post-test – Pre-test against Pre-test score; c) graph created using one-tenth bands of data to illustrate more clearly the magnitude of the dependence in b), the chained line being the theoretical scale capping effect; d) as for c) but broken down by gender; e) the dependence of the change in score between Post-test and Delayed-test on Post-test score; f) as for e) but broken down by gender.	Page 60

Figure 3.7	The dependence of change in mean PSD scores between tests in the Main Study: a) the dependence of change in PSD score Post-test – Pre-test on Pre-test PSD score; b) the dependence of the change in PSD score between Post-test and Delayed-test on Post-test PSD score; c) the dependence of change in PSD score Post-test – Pre-test on awareness of the outdoor environment.	Page 64
Figure 3.8	Distribution of standardised residuals from the regression for Post-test mean PSD scores	Page 69
Figure 3.9	Distribution of standardised residuals from the regression for Delayed-test mean PSD scores	Page 78
Figure 4.1	The hierarchical structure of variables	Page 92
Figure 4.2	A parallel, object-oriented model of student during OAA. Figure 2 from McKenzie (2003)	Page 93
Figure 6.1	Changes in pupils' mean PSD scores across tests	Page 116
Figure 6.2	The dependence of change in score between Pre-test and Post-test on Pre-test mean score for the Main Study, all	Page 118
Figure 6.3	The dependence of change in score between Post-test and Delayed-test on Post-test mean score for the Main Study, all cases, all PSD items. Graph created in Excel using one-tenth bands of data	Page 118
Figure 6.4	As for Figure 6.3 but broken down by gender. Note that each data point represents a sample of only 18	Page 119
Figure A4.1.1	A parallel, object-oriented model of student during OAA. Figure 2 from McKenzie (2003)	Page 197
Figure A4.1.2	Simple graphical representation of a learning cycle incorporating “intertwined” cognitive and affective experiential learning embedded in a linear or parallel set of learning experiences	Page 199

## List of tables

Table 1.1	Relevant effect sizes from Gillis and Speelman's 2008 meta-analysis of 44 studies	Page 7
Table 2.1	Typical 8 x half-day programmes at the three outdoor centres used in this project. Other activities that might be included are problem-solving games, camping, rafting	Page 15
Table 2.2	Project schedule	Page 16
Table 2.3	Descriptive statistics for pupils' mean scores in the Pilot Study	Page 21
Table 2.4	Measures of questionnaire reliability in the Pilot Study, for the whole questionnaire and within factors. Table above: Pre-test results; Table below: Post-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$ is Cronbach's alpha for comparison with the Spearman's rho values	Page 23
Table 2.5	Measures of questionnaire reliability in the Control Experiment, for the whole questionnaire and within factors. Table above: Pre-test results; Table below: Post-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$ is Cronbach's alpha for comparison with the Spearman's rho values	Page 27
Table 2.6a	External reliability coefficients from split-half analysis on the Control Experiment data, Pre-test:Post-test, and a comparison with the same tests on the Pilot Study data	Page 29
Table 2.6b	Spearman's rho coefficients for external reliability using the Control Experiment, Pre-test:Post-test	Page 29
Table 2.7	Measures of questionnaire reliability in the Main Study, for the whole questionnaire and within factors. Table above: Pre-test results; middleTable: Post-test results; Table below: Delayed-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$ is Cronbach's alpha for comparison with the Spearman's rho values	Page 31
Table 2.8	Descriptive statistics for pupils' mean scores in the Main Study	Page 32
Table 3.1	Results and outcomes from statistical tests on mean scores on items for both Main Study and Control Experiment	Page 43
Table 3.2	Results and outcomes from statistical tests on PSD mean scores of pupils in both the Main Study and Control Experiment. Upper table: between test; lower table: between genders	Page 51
Table 3.3	Pupils' mean scores on factor Awareness of the Outdoor Environment	Page 54

Table 3.4	Results and outcomes from statistical tests on the mean scores of pupils on Awareness of the Outdoor Environment in both the Main Study and Control Experiment. Upper table: between tests; lower table: between genders	Page 56
Table 3.5	Summary of $R^2$ values for univariate regressions between tests for pupils' mean scores within PSD factors in the Main Study and the Control Experiment	Page 66
Table 3.6	Summary of $R^2$ values for univariate regression of pupils' mean Post-test PSD scores on their mean scores in individual factors in the Pre-test, for both the Main Study and Control Experiment	Page 67
Table 4.1	Descriptive statistics for pupils mean scores in the Main Study	Page 91
Table 4.2	Main Study mean scores on all 30 questions, six Self Perception of Confidence questions and question 18	Page 95
Table 4.3	Analysis of exit questionnaire returns from pupils attending a one-week course of adventure activities at Benmore Centre	Page 100
Table 5.1	Means of teachers' scores on Low Bank Ground post-course questionnaires (N=60)	Page 109

## **Acknowledgements**

Many thanks to Pete Higgins, other staff and students of the Outdoor Education group at Moray House School of Education for making an interloper from Science and Engineering most welcome. Pete and Simon Beames have acted as sympathetic and constructive supervisors and other members of the School have provided advice and information from time to time. Headteachers and staff at the participating schools are thanked for their interest in this project and for accommodating my visits. Heads of outdoor centres involved in the project have also been extremely helpful. In particular, I would like to thank Andy Beveridge and Gill Berrow of Benmore Outdoor Centre for their support early on in the project. Special thanks to my son, Andrew, for many stimulating conversations on the theory and practice of delivering outdoor adventure activities.

# **Chapter 1            Introduction**

## **1.1      Background to this project**

At the heart of this research project is an attempt to measure the benefits that a week-long residential experience of outdoor adventure activities (hereinafter, OAA) might yield for the personal and social development (hereinafter, PSD) of children in the later years of primary school. Many, but by no means all, Scottish primary schools make a residential visit to an outdoor centre with the development of their pupil's personal and social skills in mind. The visit builds on the fact that PSD is now well established as a component of the school curriculum in Scotland, originally in the 5-14 National Curriculum (SOED, 1993) and now in A Curriculum for Excellence (LTS, 2008). Indeed, the four "capacities" of A Curriculum for Excellence – successful learners, confident individuals, responsible citizens and effective contributors – depend heavily on sound personal and social skills. The rise to prominence of PSD comes from a realisation that in the modern world, which is interpersonal, multi-cultural and secular, personal and social skills are increasingly important for the success of the individual and his/her community. A wealth of relatively recent research lies behind the positive arguments for the inclusion of PSD in the outdoor learning component of the new curriculum in Scotland, e.g. Nicol et al. (2007), although this component does not necessarily have to include OAA.

In some ways a residential week of OAA is analogous to a fieldwork visit or "the school trip", which has been a feature of school life for decades, with a steady growth in its use as an alternative learning experience complementing and enhancing classroom work. In fact, in their review of research into outdoor learning, Rickinson et al. (2004) base their report on a three-fold division of outdoor activities, which includes OAA alongside fieldwork and learning within school grounds. It is in large part their alternative, experiential nature - different place, different contents, different forms of interaction between participants – that makes such excursions memorable and an effective form of learning in both the cognitive and affective domains (e.g., Bogner, 1998; Nundy, 1999). As far as excursions that involve outdoor adventure activities are concerned, Hattie et al. (1997) identified a growth in their use beginning in the 1950s with the formation of the Outward Bound programme, which was specifically targeted at developing self effectiveness. Before that, the Scout Association had functioned since the beginning of the 20<sup>th</sup> Century to provide adventure activities and personal development opportunities. The Scottish Outdoor Education Centres were

established in the 1950s to encourage the individual to learn about themselves and their place in the world. For a comprehensive account of the development of outdoor adventure education during this period, visit Hopkins and Putnam (1993). Schools adopted this outdoor approach to personal development and throughout the second half of the 20<sup>th</sup> Century there was a growth in residential school trips involving adventure activities. However, in recent years, partly through a shift away from public funding for outdoor centres through the education service, there has been some decline (Higgins, 2002). The decline has been fuelled by increasing awareness of, and even aversion to, the risks involved in taking school children on adventure trips, as exemplified by the reaction to the Lyme Bay tragedy of 1993, and by the increased workload for teachers in other areas, in particular, assessment (Allardyce, 2008).

Despite some decline in their occurrence, there is currently strong cross-party support in political circles for a residential week of OAA for all school children (e.g., Allardyce, 2008; BBC, 2008). This support is based on a large body of research conducted in parallel with the growth in OAA that has yielded strong anecdotal and qualitative evidence for the PSD benefits children display. The work of Dismore and Bailey (2005) is a good example of this type of research with respect to primary school children, in which they used focus groups of pupils, parents and teachers and analysis of pupils' work to document benefits in both cognitive and affective learning domains and in social development. However, residential weeks of OAA are a relatively expensive school activity, and success in the competition for funds to support them is based increasingly on quantitative measures of benefit. Quoting from Allardyce (2008), "Currently, Scotland's four (sic) outdoor education centres provide training for 20,000 youngsters from schools, youth organisations and special-needs groups. Extending this to cover a one-week residential stay for 53,000 pupils a year would cost £10m". To support claims for this level of expenditure, there is a long-standing call for rigorous and robust empirical evidence that a residential experience of OAA is beneficial for the personal and social development of children (Barrett & Greenaway, 1995; Neill & Richards, 1998; Rickinson et al., 2004; Nicol et al., 2006). We can interpret this as a call for statistically-robust measures of PSD benefits, and to provide such a measure is one of the objectives of this research project.



## 1.2 OAA and PSD

If it is observed that a residential week of OAA contributes to the PSD of school children, it is important to consider why it does so. Some would argue that rather than the outcome, which might be a quantitative measure of benefit, it is the learning process that the child goes through during the week and afterwards that is important for PSD. A discussion of learning processes is presented in Appendix 4.1, relating to the interpretation of scatter in the data collected during this project. In this introductory chapter, let us now look at what confronts the primary school pupil when he/she embarks on a residential week of OAA.

For the vast majority of primary school children the residential week is the first substantial period of time they spend away from their family. Not only are they removed from this source of support, but they have to adjust to a new social structure that includes sharing a dormitory, contributing to a range of housekeeping duties and even living without television or the internet! These social parameters alone present new challenges, for some exciting and for others nerve-wracking, but for all, memorable. The location, usually a tough and remote natural environment with which they will be closely interacting, presents a unique “all-day play time” in which they are positively encouraged to have fun, develop new skills, find new challenges and take new risks. Indeed, all the primary schools I worked with in this project described the week as a “fun experience” with little or no emphasis on cognitive elements of the curriculum at this stage of the child’s development (although in some cases follow-up work contained cognitive learning elements). The adventure activities themselves, also new to the vast majority of pupils, present a set of emotional as well as physical challenges, including the elements of risk-taking just mentioned, which might instil fear, competition, and teamwork, which requires cooperation and, perhaps, leadership. These social and physical factors create a rich environment for experiential learning and shape the responses of the pupils. Throughout, pupils’ communication skills are challenged both from a learning and expressive point of view. The role of the pupils’ teacher changes, essentially allowing the child more freedom to develop, and the role of the group instructor as a facilitator, *inter alia*, encourages independent development. There is, therefore, a plethora of new and challenging situations to provide learning opportunities. Given this learning environment, it is entirely appropriate that theories of learning such as Kolb’s Experiential Learning Cycle (Kolb, 1984) and Walsh & Golins’ State of Adaptive Dissonance (Walsh & Golins, 1976) have been applied to the personal and social learning process.

During a residential week of OAA there is a plethora of new and challenging situations to provide learning opportunities. Numerous recent publications have addressed the question of how adventure programme outcomes are achieved (McKenzie, 2000; Dismore & Bailey, 2005), but, according to Dismore and Bailey, "... few of these reports provide, or claim to provide, systematic analysis of the processes and outcomes of OAA", perhaps because they have preferred to address the question in terms of the components of the experience, described in the previous paragraph, rather than in terms of theories of learning. Rickinson et al. (2004) also conclude that the nature of learning in outdoor education is a research "blind spot". It seems odd that this is the case, given the encouraging start from models proposed in the 1970s and 1980s. Nevertheless, it appears that there is a need for rigorous, empirically-based research into the learning processes at work during outdoor education, including outdoor adventure activities. Thus, another objective of this project is for the author to better understand the processes of learning associated with a residential week of OAA. Later in this thesis we consider the importance and implications of the results for understanding the likely PSD learning processes.

### **1.3 The search for a quantitative measure.**

Although semi-quantitative, qualitative and anecdotal evidence of the benefits of OAA for PSD abounds (Cooper, 2004; Dismore & Bailey, 2005; Rickinson et al., 2004), numerous quantitative studies aimed at securing evidence have still to deliver outdoor educationalists in the UK, and Scotland in particular, with the proof that these weeks are educationally meaningful and worthy of an effective level of funding. In their 1994 meta-analysis of quantitative research on young people 11 -18 years old over the previous 25 years, Cason & Gillis (1994) said, "While many practitioners intuitively "know" that adventure programming is effective for adolescents, they are often at a loss when asked for some [statistical] research data ... to support their claim". Thus, they argued their case for conducting a meta-analysis to establish a robust measure of benefit. Although it is not clear whether any of the studies they used came from the UK, their comments could be applied to the situation here. Referring to the UK specifically, Rickinson et al. (2004), reviewing research published on outdoor learning in the 1993 – 2003 period, concluded that "there is a particular need for more UK-based research", and from a quantitative perspective drew heavily in their review on the meta-analysis of Cason and Gillis (1994) and another from overseas by Hattie et al. (1997). Rickinson et al. go on to identify a trend away from quantitative research in the 2000s and the emergence of a greater number of qualitative

studies in the wake of concerns about methodological rigour. The positivist research community seemed to feel that it had gone as far as it could in trying to quantify the benefits of OAA to young people, in that these benefits were either demonstrable or not. This point is revisited in Chapter 4, noting the implications of the results from this study for a continued way forward for the positivist research approach.

### 1.3.1 Effect size

Since effect size is used later in this thesis as a measure of the benefit of a residential week of OAA for PSD, it is useful to review effect sizes reported in published research. Cohen (1988) defines effect size as “the degree to which the phenomenon is present in the population” and expresses it mathematically as,

$$d = \frac{m_A - m_B}{s}.$$

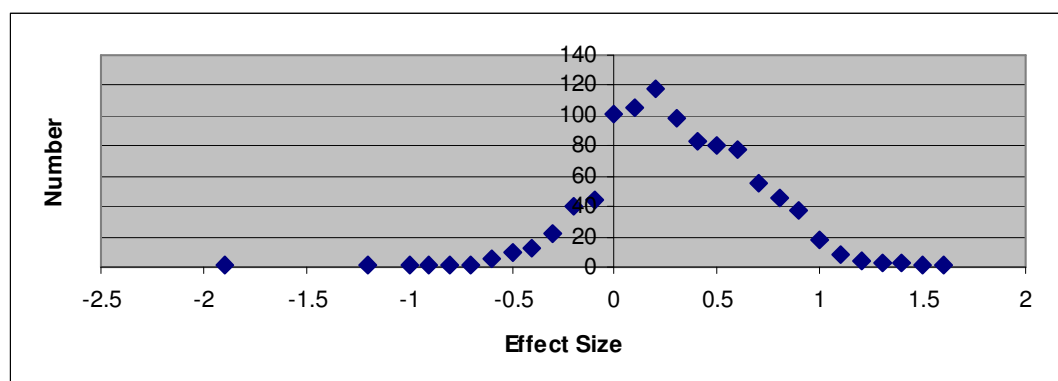
In words relevant to this study, it is the difference in the mean scores of two administrations of a questionnaire, before and after an intervention, or the difference in the mean scores between the intervention group and a control group, or between boys and girls, divided by a standard deviation from the data, sometimes from the later of two administrations, sometimes the mean. Working definitions of the magnitude of effect size are “small” if  $d=0.2$ , “medium” if  $d=0.5$  and “large” if  $d=0.8$  (Cohen, 1988). However, in due course it will be seen that the effect sizes from analyses in this thesis are given as Pearson correlation coefficients,  $r$ , derived according to Field (2009), not as Cohen’s  $d$ . Cohen gives the relationship between the two types of effect size as,

$$r = \frac{d}{\sqrt{d^2 + 4}}$$

from which it can be calculated that for the  $d$  values just quoted,  $r=0.1$  when  $d=0.2$  (small),  $r=0.24$  when  $d=0.5$  (medium), and  $r=0.37$  when  $d=0.8$  (large). It is understood that in educational psychology,  $r=0.37$  would not be considered “large” (C.Iannelli, personal communication), therefore, in the analyses presented here, the  $r$  values are interpreted as small if  $r=0.1-0.2$ , medium if  $r=0.3$  and large if  $r=0.5$ . These are, in fact, the values suggested by Cohen (1988) for research in educational psychology and adopted by Field (2009). In the following discussion of published effect sizes, it appears that all the effect sizes are Cohen’s  $d$ , although not all publications say that they are.

In so far as effect sizes reported in the meta-analysis by Cason & Gillis (1994) can be isolated for children of late primary school age on a one-week residential visit of adventure programming, they cluster around 0.14. This would be a measure of improvement in self esteem over the course of the week. Similarly, for Hans' (2000) meta-analysis, in so far as the relevant results can be isolated, effect sizes for greater internalisation of locus of control range from 0.3 to 1.2. Extracting effect sizes relevant to this research from the meta-analysis of Hattie et al. (1997) is equally tricky. Using participants from 11 years old to adult the authors were able to say that "the effects were not moderated by age". From my reading of the paper, values for effect size on "short programmes" of around 0.2-0.3 would seem to be appropriate for students of "normal" background. Figure 1.1 shows the distribution of all 1728 effect sizes extracted from 96 studies (none within the UK) by Hattie et al., for which the mean is 0.34. In both meta-analyses, Cason & Gillis and Hattie et al., the overall mean effect size for all categories of participants on all programmes is slightly higher than the values extracted here for young people closest to the late primary school age of interest – 0.31 as opposed to 0.14 for Cason & Gillis and 0.34 as opposed to 0.2-0.3 for Hattie et al. – suggesting that of all groups, normal children register relatively low benefits. These relatively low measures of benefit might be real, although that would conflict with the strong qualitative and anecdotal evidence for benefit, or they might reflect difficulties we have in making quantitative measurements of benefit with children. Possible weaknesses in the quantitative method are explored later in this thesis. In an interesting additional result of Hattie et al.'s research it was found that short-term positive effects were followed by

**Figure 1.1** Distribution of all effect sizes extracted by Hattie et al. (1997) from 96 studies. These are for all age groups on all programmes. Mean = 0.34. N.B. many negative effects.



additional gains of mean effect size 0.17 in follow-up assessments. A longitudinal study is part of the research reported in this thesis with which this result can be compared.

Some research targeted at specific adventure activities does allow easier extraction of results for late primary age children, albeit still based on data collected overseas. An example is the use of ropes courses for personal development, for which Gillis & Speelman (2008) carried out a meta-analysis of 44 studies. Ropes courses are commonly a component of the generic, taster OAA weeks offered by Scottish outdoor centres. The results are shown in Table 1.1.

**Table 1.1** Relevant effect sizes from Gillis and Speelman's 2008 meta-analysis of 44 studies.

Subset of the 44 studies	Mean effect size
Children 11 - 15 years old (8 studies)	0.46
"Developmental" studies across all ages (22 studies)	0.47
"Developmental" studies, children 11-15 years old (2 studies)	1.33
Post-tests (43 studies)	0.45
Delayed-tests (12 studies)	0.23 (i.e. -0.18 from Post-tests)

There is some evidence here of higher effect sizes than those in the more general meta-analyses, but, on the other hand, a clear loss of benefit is reported in follow-up assessments, -0.18 as opposed to the +0.17 reported by Hattie et al. (1997).

Rickinson et al. (2004) include a comprehensive review of the impacts of outdoor adventure activities, reviewing several semi-quantitative studies from the UK, but none in which effect size is calculated. They conclude from these that "There is substantial research evidence to suggest that outdoor adventure programmes can impact positively on young people's: attitudes, beliefs and perceptions, ... interpersonal and social skills." However, they found less strong evidence for cognitive and physical benefits from OAA, except where the focus of the intervention is in such areas, such as the "fieldwork effect" of Nundy (1999). Nundy's research, which was with 10 and 11 year old school children undertaking residential geography fieldwork in southern England, is very interesting, because it seeks to not only establish the cognitive benefits of the fieldwork but also the affective benefits in both academic discipline and personal/social (PSD) areas. The results indicate that the children acquired affective benefits to the extent of effect size 0.21 in the academic discipline area but only 0.08 in the personal/social area. On the basis of the value of only 0.08, we can say that residential fieldwork does not yield social/interpersonal benefits to the same extent as

adventure activities weeks, as revealed by the meta-analyses described above. It does, however, illustrate the fact that while individual quantitative studies yield varying amounts of benefit, most commonly effect sizes of about 0.2 are found. Nundy's research is revisited in the Appendix to Chapter 4 in connection with a "fieldwork effect" from OAA leading to accelerated learning processes in personal and social skills.

In the search for quantitative measures of benefit in the peer-reviewed literature, results from the UK are few, and from Scotland non-existent. A few results can be found in postgraduate theses, however. Christie (2004), researching into the Raising Achievement Project of North Lanarkshire Education Department, used the 1997 version of the Life Effectiveness Questionnaire (LEQ) of Neill (Neill, 2007) to assess the PSD benefits for 14-16 year old pupils participating in a week-long Outward Bound programme relative to a control group of non-participants. She found "no significant difference between the two groups in terms of the LEQ scores", that is, no statistically significant effect size. And yet, in the interviews with participants and through her own observations she found "positive overall effects in terms of the students' perception of their social ... skills". This, then, is a perfect illustration of a current problem: that semi-quantitative, qualitative and anecdotal evidence of the benefits of OAA for PSD abounds, but quantitative studies aimed at securing evidence have still to deliver a consistent measure of benefit.

Why is it difficult to obtain consistent quantitative measures of the benefits of OAA for PSD? Many researchers blame it on poor or inconsistent research methodology. In fact, Cason and Gillis (1994) include research design quality as one of the variables in their meta-analysis, with the conclusion that poorer design yielded larger effect sizes. Further contributing factors might be the large number of variables on which the quality of the OAA experience depends and the adequacy of the measuring instrument, most probably a questionnaire. At a more fundamental level there is concern over whether PSD is measurable at all. Conscious of the fact that published league tables of examination results will encourage schools to focus on what is measurable in the curriculum, the book edited by Inman et al. (1998) explores extensively the question of whether a PSD measure can be developed that is comparable to the measures for academic subjects. Unfortunately, the authors do not seem aware of several decades of research in which attempts have been made to measure PSD quantitatively using the medium of OAA. In Chapter 4 we return to the issue of the dichotomy between measures of benefit for PSD from qualitative and quantitative approaches to research.

## 1.4 The nature of this study

In the preceding sections of this Introduction several aspects, indeed issues, of the provision of OAA for primary school children and the measurement of the benefit of these activities for PSD have been discussed. From these issues emerged the overarching aim of this research, which is to attempt to gather some quantitative evidence in support of the cross-party call in the Scottish Parliament for a compulsory residential experience of outdoor adventure activities for all school children (Allardyce, 2008; BBC, 2008). In turn, to meet this aim, five research objectives were identified:

- To design and test a new questionnaire for use with late primary schoolchildren to make quantitative measures of changes in their perception of their own personal and social skills in association with a one-week visit to an outdoor centre,
- To carry out an experiment to make measures just before, just after and about 2-3 months after the pupils' residential experience of OAA,
- To analyse the data to identify changes in scores that might indicate benefit, and test these for statistical significance,
- To further analyse the data to discover whether there are certain attributes amongst children of late primary age – environmental awareness, personal and social awareness, etc. – that make them particularly amenable to benefit, and
- To consider the implications of the results for researchers and practitioners, including policy-makers and funders of residential visits, and understand them in the context of theories and models of learning that might apply during OAA.

In most respects this research is an exploration of what might be obtained by way of robust empirical support for funding for all Scottish schoolchildren to experience a residential week of OAA. It is clear that many education policy makers are already convinced of the benefits of residential OAA, but there are barriers to making “residential” a compulsory component of the curriculum. These barriers might be the necessary infrastructure and running costs, safety concerns, teacher perception of risk and teacher engagement elsewhere, specifically with achieving the targets for academic league tables. It is hoped a substantial, robust,

quantitative measure of positive benefit for PSD will be observed, but it is not assumed or hypothesised that a positive measure specifically will be observed.

It is important to note that because it is not assumed or hypothesised that a positive measure of benefit will be observed, throughout this work 2-tailed statistical significances are calculated and reported. In Section 1.3 it was reported that both positive and negative measures are possible, although positive are more common than negative. As it happens, positive measures are obtained at one stage and negative measures obtained at another stage of this project, thus justifying the cautious approach. In some specific instances, a 1-tailed significance might be justified in view of trends in the values of mean scores, but for consistency of presentation 2-tailed significance is adhered to. These instances do not affect any of the main conclusions.

This research also incorporates a measure of the effect of pupils' environmental awareness on their ability to benefit from a residential week in an outdoor setting. This aspect of outdoor learning received quite a lot of attention from Rickinson et al. (2004) in their review. Much of the reviewed literature was inconclusive, through the use of small samples or not being relevant to primary school children, but on balance it was said that "the evidence of a positive link between outdoor adventurous activities and environmental understanding and values is not strong". The fourth objective here is to test this claim.

I am measuring the PSD benefits of OAA at the late primary school level essentially for opportunistic reasons. This is the time in their education when school pupils are commonly taken on a residential visit to an outdoor centre. There are two reasons for schools choosing this time. The most relevant reason for the purposes of this project is that pupils are learning about personal and social development and are sufficiently cognisant with the concepts to understand what a profound impact a week spent away from home undertaking challenging social and adventure activities might have on their knowledge of themselves. In effect, schools are using the residential week as a PSD field trip to reinforce what is being learnt in the classroom. The second reason is that the residential visit provides the opportunity for pupils from a cluster of primary schools to meet together and make new friends before they make the transition to their high school.



## **1.5 Structure of this thesis**

Following this Introduction, Chapter 2 describes experimental design, the project schedule and the development of the questionnaire. Reliability testing of the questionnaire showed it to be only marginally reliable, with changes in reliability across various applications. However, by appealing to a corollary of the assumptions behind the test-retest approach to reliability testing, some evidence emerges for pupils improved understanding of the questionnaire over the two weeks between Pre-test and Post-test. The work in this Chapter might be of interest to researchers in project methodology and to teachers.

Chapter 3 is the main data analysis chapter in which pupils' mean scores on variables are presented, differences calculated and inter-dependences between variables established. Using gender as a variable, it is apparent that girls score higher than boys throughout the study; and using low-scorers and high-scorers as two groups, it seems that low-scorers gain more benefit. For practical purposes in the time available, it became necessary to work with two different groups of schools. It had not been the intention to have school group as a variable, but during data analysis it became apparent that the groups were performing differently. This difference is interpreted in terms of the socio-economic background of the schools. These and other results, will be of interest to researchers, teachers and outdoor practitioners.

Chapters 4 and 5 reflect on the implications of the project results for researchers and practitioners respectively. It is a pity that the quality of this study has been compromised by the marginal reliability of the questionnaire, although the internal consistency of all the results leads to the belief that we observe at least an indication of the positive benefit of OAA for the PSD of late primary school children. With the knowledge and understanding of research in education I have gained from undertaking this work, a much more robust study could be undertaken. Chapter 6 provides a summary of the whole project.

## **Chapter 2                      Experimental and Questionnaire Design**

### **2.1      Introduction**

This project sets out to make quantitative measures of the benefits of a residential week of outdoor adventure activities (OAA) for the personal and social development (PSD) of late primary school children. It has been argued that although it is very difficult to assess or measure PSD it is not impossible (Inman, Buck, & Burke, 1998), and numerous attempts have been made within outdoor education to do this, with some success. Typically with OAA, in the more generic studies of a mix of PSD attributes of pupils attending courses with a mix of activities, small benefits (effect sizes) have been measured, whereas with more targeted groups, such as young offenders, or with courses of a more specialised nature, such as high ropes, larger changes have been detected. A few meta-analyses summarise these results - see Neill and Richards (1998) for a review of the meta-analyses.

Despite disparaging comments about the likely success of attempting to measure what is, after all, a social phenomenon, and the prospect of many different variables coming to bear on the outcome, it was decided to undertake a study that is as generic as possible. Greig et al. (2007) discussed the relative merits of quantitative and qualitative research with children and the precautions that should be taken with quantitative research. They emphasised the need for representative samples and the benefits of mixed-method research. The laws of the natural sciences permit scientific experiments to yield meaningful results with just a few high-quality observations, but even if we assume that human beings ultimately function according to the laws of nature (e.g. biochemical processes in the brain or genetic bias), the complexity introduced by a multitude of different personal experiences (nurture) that are stored in memory defeats a purely scientific approach to quantitative research with human beings as social entities. In this research project the principal objective was to try to study the impact of OAA on the PSD of the average 10-12 year old child by collecting quantitative data from a representative sample of ordinary late primary school pupils and analyse this sample statistically. The statistical significance of the outcomes provides a measure of the probability of the impact occurring in the whole population. In the course of doing this it was necessary to make assumptions and identify the constants and variables within the study to ensure such terms as “average”, “representative” and “ordinary” do apply. The variables are described below.

Although this research does not take a qualitative approach of any substance, it is acknowledged that a mixed-methods approach to research with children might yield more useful results. However, there has been a remarkable dichotomy between the results of qualitative and quantitative research on the PSD benefits of residential outdoor adventure activities that asks some questions of a mixed-methods approach in this area of research. Qualitative evidence from interviews with teachers, parents and pupils, and anecdotal evidence from various sources, indicates strongly that there is great benefit to be gained, but quantitative measures are ambivalent and frequently give small estimates of benefit, even in the same study. A typical example of this is the research of Christie (2004) with secondary school pupils from North Lanarkshire who visited an Outward Bound centre for one week. She found that, "...interviews with those who participated in the programme pointed to positive overall effects in terms of the students' perception of their social .. skills.", whereas "...the quantitative study showed no significant difference between the two groups [those who participated in the programme and those who didn't] in terms of the LEQ [Life Effectiveness Questionnaire] scores." In situations like this the applicability of the questionnaire might be doubted. In part, it was this kind of situation that encouraged me to design my own questionnaire. Mixed methods were used to test the viability of the new questionnaire, however.

## **2.2 Project structure**

Background reading in order to formulate the project took place during the summer of 2008. The initial plan was to use quantitative data to train an artificial neural network that would be able to predict the benefits accruing to pupils with particular prior PSD skills, and thus glean something about the learning processes at work amongst the pupils during OAA. It very quickly became clear that this would be too difficult a task within only two years of research, although, in theory, it is something that could be attempted in the long term if a number of sufficiently reliable data sets become available to train the network. The objective was, therefore, scaled down, with a revised plan to use the data to test a structural equation model linking predictors (prior PSD skills) and outcomes (benefits from a residential week of OAA). Research started in earnest with this objective in mind, but in the event, the non-parametric distribution of the datasets restricted the analysis of links between predictors and outcomes to simple regression modelling of mean scores.

### 2.2.1 Variables, constants and assumptions

Irrespective of the approach towards data analysis and modelling, decisions had to be made on which parameters in the study would be constants and which would be variables. The decision to undertake a generic study of the PSD of pupils attending their first, taster course of a mix of outdoor adventure activities had the potential to generate many variables. The decisions can be summarised as follows.

- Measures of personal and social skills obviously had to be variables. The construction of these variables is described in the section on questionnaire development, which follows.
- A further variable was introduced into the questionnaire, called “Awareness of the Outdoor Environment”, to measure pupils’ familiarity with and enjoyment of the outdoor environment, which might influence their susceptibility to benefit from a week in the countryside.
- Gender was treated as a variable during data analysis.
- Age was treated as a constant, since the focus was on Primary 6 and Primary 7 pupils of 10-12 years old, although much of the scatter in data collected might be attributable to the wide spread of social competencies within this age range.
- Initially, it was assumed that the two groups of schools contributing to the project functioned as a constant, but it became clear that the two groups used for different parts of the project generated different scores on the questionnaire; the two groups of schools became a variable, therefore.
- The philosophical approach of the schools to the residential weeks was also taken to be a constant. All the schools involved treated the week as a fun week, with an emphasis, if any, on “life experiences” (PSD). There was no academic work apart from the completion each evening of a pupil’s daily diary. All the teachers involved with the visits said they did little or no preparatory or follow-up classwork related to the residential week.
- Although three different outdoor centres hosted the schools, their programmes and the competence of their instructors were considered to be constant. All three outdoor centres deliver similar introductory programmes of OAA over 4 days, Monday lunchtime to Friday lunchtime (Table 2.1). The countryside location (the “place”) of the centres was also considered to be a constant.

**Table 2.1.** Typical 8 x half-day programmes at the three outdoor centres used in this project. Other activities that might be included are problem-solving games, camping, rafting.

Activity	Benmore Centre, Argyll	Lagganlia Centre, Speyside	Dounans Centre, Perthshire*
Abseiling	√	√	
Archery			√
Climbing	√	√	√
Gorge walking	√		√
Kayaking/canoeing	√		√
Mountain biking	√	√	√
Mountain day		√	
Orienteering		√	√
Ropes course	√	√	√
Sailing	√		
Ski slope		√	
Woodland/forest walk	√	√	√

\* Some school visits had to be relocated to Broomlee, Scottish Borders.

In Chapter 3, on data analysis, scores on the variables within the questionnaire that was used were averaged into pupil means. The means were used for the comparative study of variables and for the way in which gender or school group influenced them.

## 2.2.2 Project schedule, the sample and data collection

An important aspect of the project structure, which had a bearing on what could be achieved in the research, was the project schedule. The main constraints on the schedule were the time at which data collection could take place - which was the time at which Lothian schools were undertaking their residential visits - and the need to build up a statistically useful sample size. Table 2.2 outlines the main phases of the project. Statistical analysis of the data was ongoing throughout the project and the time devoted to specific phases indicated by the timelines in the Table.

Initially, a control experiment, something akin to a randomised controlled trial, was not planned for this project, but after a Pilot Study to test the reliability of the questionnaire designed specifically for this project (see Section 2.3), in which the reliability appeared to improve in the test following the pupils' week at Benmore Centre, it was felt necessary to conduct a test-retest of the questionnaire reliability. Thus, a Control Experiment came about. This served both as a test of questionnaire reliability and as a comparison for the results of the Main Study.

**Table 2.2.** Project schedule. Data analysis was taking place throughout 2009 and 2010.

Phase	A 2008	W 2009	Sp 2009	Su 2009	A 2009	W 2010	Sp 2010	Su 2010	A 2010
Preparation and familiarisation	X	X							
Questionnaire design and testing		X	X	X			X		X
Questionnaire Pilot Study			X	X					
Control Experiment						X	X		
Main Study Pre-tests					X	X	X		
Main Study Post-tests					X	X	X		
Main Study Delayed-tests							X	X	
Thesis writing								X	X

A=Autumn; W=Winter; Sp=Spring; Su=Summer

It was decided that a sample size of at least 200-300 would be desirable for the project. This was based on Cohen's (1992) recommendation of a statistical power of 0.8, and knowing that even if only a small benefit was measured its effect size would be about 0.2 and that the default probability of an effect being statistically significant is normally taken to be 0.05 or smaller. Data collection was carried out with two groups of schools. This was not ideal for the project because it introduced another variable (see above) but was necessary because of the project schedule and the timing of the schools' visits to the outdoor centres. However, the schools that participated were used simply because they happened to be attending one of the three outdoor centres at the time I wanted to collect data. This made the sample of Primary 6 and 7 pupils to all intents and purposes a random one, although, strictly speaking, pupils from the same school are not independent entities. The primary schools used for the Pilot Study of the questionnaire and the Control Experiment were: Castlevue, Leith and Longstone, all Edinburgh; Hopefield, Bonnyrigg; and Strathesk, Penicuik. This group attended Benmore Outdoor Centre and yielded a sample size of about 100 on each occasion. The schools used for the Main Study were: Royal High, Edinburgh; Dunbar Primary; Linlithgow and Linlithgow Bridge Primaries; Beancross, Bowhouse and Moray, Grangemouth; and Maddiston and Wallacestone, Falkirk. This group attended Lagganlia or Dounans Outdoor Centre and yielded a sample size of about 350. All schools were from the state sector.

To engage the schools in the project, first of all I asked the outdoor centres to provide me with a list of schools using their centre in the timeframe available for data collection. All the centre heads are personal friends or acquaintances who immediately appreciated the merits of the project. The Headteacher of each school was approached for permission to use their school in the project, including making available to them a recommendation from the appropriate centre head, the project description, the ethical clearance from Edinburgh University and my CRB clearance (as a qualified sports coach). Permission was granted by all but one school approached in this way, with considerable interest shown in the project. Thereafter, contact with the pupils was *via* their class teachers. Each participating school and outdoor centre has now received a summary report of the results of the project extracted from the contents of this thesis.

In the case of the Pilot Study, the pupils were tested in the week before they attended Benmore Centre and in the week afterwards. In the case of the Control Experiment, testing was carried out about two months before the pupils went on their week of OAA at Benmore with no indication to the pupils about the significance of the questionnaire. The classes involved completed the questionnaire two times, about two weeks apart. In the case of the Main Study, the pupils were tested in the week before they attended Lagganlia or Dounans and in the week afterwards, and tested again 2 or 3 months later. Throughout this thesis the first test is called the Pre-test and the second test is called the Post-test; the longitudinal study uses the Delayed-test. Some small adjustments were made to the questionnaire following the Pilot Study in 2009, otherwise the questionnaire was the same throughout. Although the 30 questions were given in a different order in each test, the class teachers and my research colleagues were quite confident that pupils would not remember specific questions from one test to the next.

To administer the questionnaire I went into each school by arrangement with the class teacher. I noticed that teachers introduced me and the questionnaire with an emphasis on participating in a research project, a concept that pupils were becoming familiar with through their own work, thus gaining their attention and sense of value of the exercise. The pupils took about 10 minutes to complete the questionnaire, working completely independently so that the response from each pupil was theirs alone. In order that returns could be matched up across tests, pupils were asked to write their initials and gender on the cover sheet. Apart from these essential indicators of identity, I had no knowledge of any individual pupil. This approach to pupil anonymity was found to be perfectly acceptable to the schools.

## **2.3 Questionnaire development**

### **2.3.1. Rationale**

A questionnaire to make the necessary measurements was developed specifically for this project. This might seem like an unnecessary extra task, since there are several questionnaires in the literature that are designed to measure personal and social skills and their development and, furthermore, are widely used, e.g. the Life Effectiveness Questionnaire (LEQ) (Neill, 2007). However, there are few, if any, questionnaires freely available and designed specifically for use with 10-12 year old children – the LEQ is primarily for adolescents. Children develop at such a rapid rate, especially in the transition from primary to secondary school when they are developing as independent learners, that to work with them at this stage is to be capturing a moment in time in their rapid development. Their learning curve is so steep that just a small difference in their age can mean a large difference in their understanding of the research project. For survey purposes, this means that a tailored questionnaire and large samples from which scores can be averaged, are particularly important.

Although children at 10-12 years old are developing the ability to think in more abstract terms, it was decided to structure “questions” in the questionnaire as statements about the child him/herself. These were personalised statements of the form “I am ...” or “I believe ...”, which the individual can readily associate with from their own experiences. Moreover, following advice in publications such as Borgers et al. (2004), the statements were made simple, unambiguous and as short as possible. On the other hand, there is a “directionality” to these questions that could potentially encourage pupils to respond with high scores. Indeed, the questions were all phrased in accordance with the hypothesis that high scores in one area would parallel high scores in other areas. This led to the datasets having negatively skewed distributions of scores, but the key point, that there was a distribution of scores that would discriminate between high-scoring and low-scoring pupils was satisfied. The first version of the questionnaire contained four reverse-meaning questions, but class teachers and their pupils advised against the use of these during an initial validity check. This advice was confirmed on examining the scores from the Pilot Study, in which the 4 questions yielded mixed and unreliable responses. These four questions were changed into forward-meaning questions to be like all the other questions.



It was considered impossible to measure personal and social skills directly because it is not possible to ask children of 10-12 years old the straightforward question, “What is the level of your personal and social development?” Instead, PSD was represented by four factors, “Self-Perception of Competence”, “Self-Perception of Confidence”, “Awareness of Others” and “Attitude towards Groups”, constituting two personal and two social factors, each quantified by the pupil’s mean score on six questions in the questionnaire. The questions were phrased so that a child of 10-12 years old would be able to understand straightaway what was being asked. The topics addressed by the questions were inspired by several existing questionnaires, many of them published. The PSD topics were also informed by the content and phraseology used in the Scottish National 5-14 Curriculum guidelines for personal and social development in late primary years (SOED, 1993) and in the experiences and outcomes of the Health and Wellbeing area of the new Curriculum for Excellence (LearningandTeachingScotland, 2008). It is fair to say that the groups of six questions were less focused than they are in other questionnaires, e.g. LEQ, with potential impact on the reliability of the questionnaire. On the other hand, pupils’ views of their PSD skills in a wide range of situations would contribute to mean scores. This was in keeping with the intended generic nature of this study.

A fifth factor was included, “Awareness of the Outdoor Environment”, also of six questions, to test the hypothesis that greater awareness of the outdoors would lead to greater benefits from OAA.

Again following the research of Borgers et al. (2004), a short, four-point Likert scale was used as an interval scale for scoring the questionnaire. Although their research did not differentiate between children in the age range 8 to 16 years old, their considered conclusion that a four-point scale is the best choice for a reliable questionnaire would seem to be appropriate for 10-12 year olds. The wording for the four choices on the scale was also tested on teachers and pupils during an initial validity check and found to be clearly understood.

It is worth noting that in December 2009, Mygind published a questionnaire for children of primary school age that is structured very much like mine, but this was, of course, too late to influence my own design (Mygind, 2009).

The final questionnaire is in Appendix 2.1. The question order is the one used for the Pre-test administrations and comprises one question from each of the five factors for questions 1 to 5, then another question from each of the five factors for questions 6 to 10, and so on. The question order was shuffled for other administrations. During the Control Experiment, mention of outdoor activities on the frontispiece was removed.

### **2.3.2 Questionnaire validity and reliability**

The validity of the questionnaire, essentially understanding and interpreting its contents correctly, was considered to be particularly important for the purposes of working with primary school children, who are still developing their understanding of the world at a rapid rate. On the other hand, it was expected that reliability might not be as strong as it would be for questionnaires used with older people who have developed consistent, even fixed, views on issues, and are perhaps more consistent in their understanding of questions. Also, it was expected that the more wide-ranging nature of the questions with each factor would affect reliability.

#### **Initial validity**

The first draft of the questionnaire contained 52 items with the intention of reducing these to about 30 on the basis of critical validity tests. Validity was tested with the first draft by obtaining feedback from departmental colleagues and then taking the questionnaire to Benmore Centre to ask a party of 65 11-12 year old pupils and their five teachers from Edinburgh and Lothian primary schools and the Centre staff for their opinions. I was particularly interested in which of the 52 questions were best understood by the pupils at the same time as addressing the aims of this project and the concepts that the pupils had been exposed to in the curriculum. These were tests for content validity and face validity respectively. The feedback received included a recommendation to avoid using reverse-meaning questions but approval of the phraseology in the four-point Likert scale. On the basis of this feedback a second draft of the questionnaire with 30 items was written and then tested quantitatively at Benmore Centre in a Pilot Study. At this stage, four reverse-meaning questions were still in the questionnaire to satisfy my curiosity as to pupils' responses.

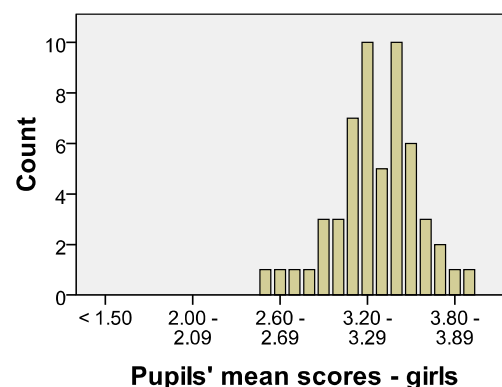
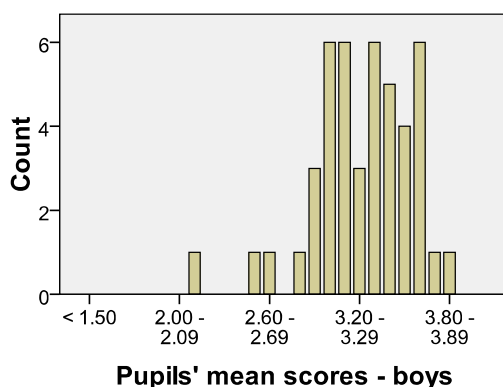
The Pilot Study used the questionnaire in a Pre-test and a Post-test in the week before the pupils visited Benmore and the week afterwards respectively. 100 pupils were sampled. No

statistical tests for validity were carried out on these data, partly because there were no external measures or criteria with which my questionnaire could be compared. However, the teachers consulted at Benmore did advise me that girls would give more socially-acceptable scores (higher, in view of the nature of the questions) and be less self-critical (less scattered scores) than the boys. This is investigated and confirmed in Table 2.3, which indicates that if analysed by gender, the mean scores for girls are slightly higher and the standard deviations (square-root of variance) are slightly smaller. The histograms accompanying this

**Table 2.3** Descriptive statistics for pupils' mean scores in the Pilot Study.

	Pre-test mean	Pre-test standard deviation	Post-test mean	Post-test standard deviation
<b>Boys</b>	3.25	0.34	3.26	0.37
<b>Girls</b>	3.30	0.28	3.36	0.30

**Distribution of pupils' mean scores, Pilot Study Pre-test**



**Distribution of pupils' mean scores, Pilot Study Post-test**

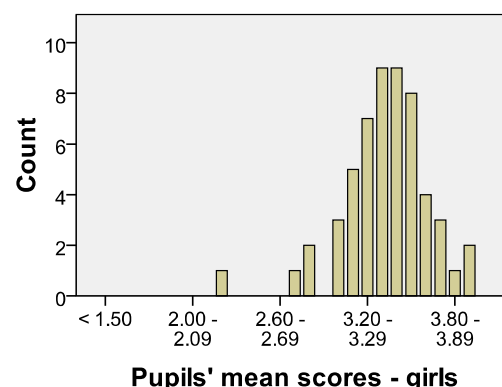
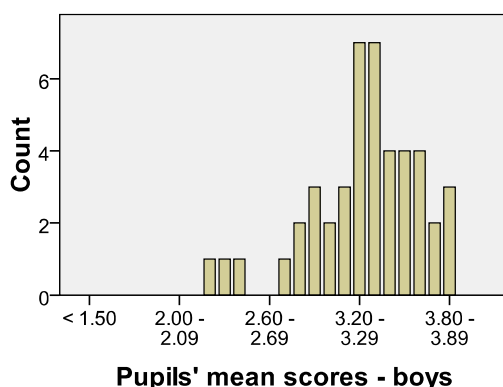
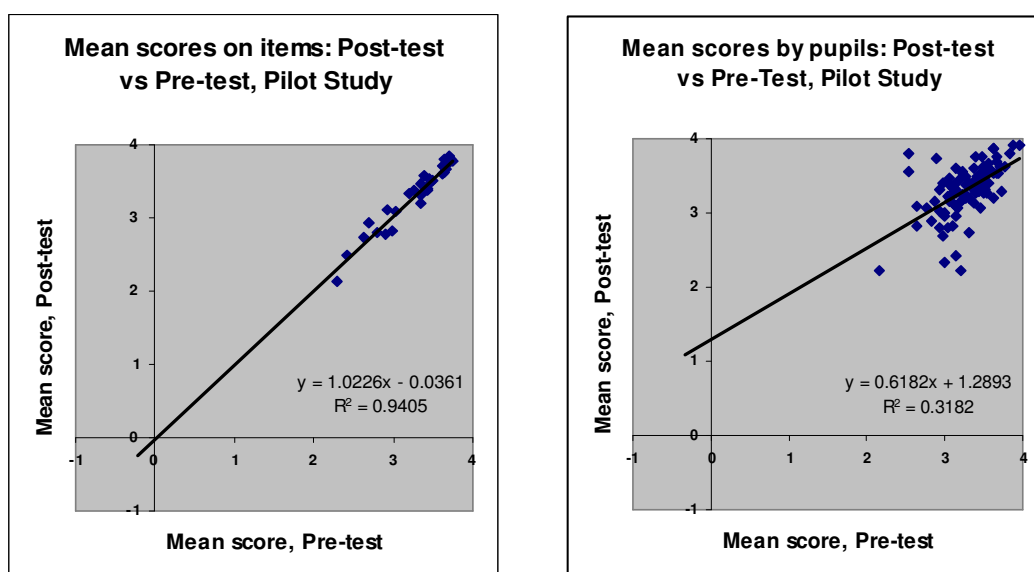


table show that pupils' mean scores on the 30-item questionnaire have a tendency towards non-normal distribution but are not so non-normal in their distribution as to make these measures unreliable. The low scores from the boys turned out to be typical of this project and, in my opinion, reflect their more self-critical nature.

### Pilot Study reliability and its implications

Data from the Pilot Study provided the opportunity to make the first tests of questionnaire reliability. Tests for internal reliability were carried out on the Pre-test and Post-test data. Because there was an intervention between the two tests this was not a test-retest experiment in which external reliability could be tested, but the Control experiment conducted at a later date did provide such an experiment, and is described shortly. In this context, it is worth noting that in the Pilot Study there is a strong linear relationship between mean scores on items in the Pre-test and Post-test (Figure 2.1, Left), indicating that irrespective of the changes in pupils' scores between tests (Figure 2.1, Right) individual questions attract similar mean scores. A similar relationship is seen in all experiments and for all five of the individual factors in the questionnaire. It is also worth noting that with the data from the Pilot Study being only marginally non-normal in their distribution, it is likely that parametric and non-parametric tests for reliability will give similar answers provided that the Likert scale is approximately linear.

**Figure 2.1** Left: strong linear relationship between mean scores on 30 questionnaire items, Pilot Study. Right: similar graph for mean scores of the 100 pupils.



Non-parametric tests for reliability were carried out using the Spearman's rho rank order calculation for mean scores on items. Tests were carried out to obtain correlation coefficients between items, and between items and the total score, for both the whole questionnaire and for the five individual factors. The coefficients are summarised in Table 2.4.

**Table 2.4** Measures of questionnaire reliability in the Pilot Study, for the whole questionnaire and within factors. Table above, Pre-test results; table below, Post-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$  is Cronbach's alpha, for comparison with the Spearman's rho values.

Items tested	Pre-test, item:item					Pre-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.13	0.14	0.46	32	0.84	0.34	0.36	0.56	73
Awareness of Outdoor Environment	0.17	0.17	0.31	33	0.51	0.52	0.53	0.64	100
Perception of Competence	0.19	0.22	0.35	53	0.53	0.54	0.54	0.59	100
Perception of Confidence	0.17	0.13	0.34	40	0.52	0.53	0.53	0.60	100
Awareness of Others	0.16	0.19	0.35	46	0.41	0.51	0.52	0.57	100
Attitude towards Groups	0.07	0.06	0.24	13	0.38	0.44	0.41	0.63	100

Items tested	Post-test, item:item					Post-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.16	0.18	0.53	44	0.79	0.40	0.42	0.64	83
Awareness of Outdoor Environment	0.13	0.13	0.37	33	0.39	0.49	0.53	0.62	100
Perception of Competence	0.24	0.23	0.42	74	0.58	0.57	0.55	0.68	100
Perception of Confidence	0.26	0.25	0.53	53	0.70	0.59	0.62	0.71	100
Awareness of Others	0.23	0.22	0.45	60	0.63	0.57	0.56	0.68	100
Attitude towards Groups	0.29	0.20	0.49	47	0.59	0.53	0.55	0.65	100

Garson (2009) suggested that as a yardstick for adequate reliability a median rho value of  $\geq 0.6$  might be assumed. If Cronbach's alpha is used as a measure,  $\geq 0.7$  might be assumed (Loewenthal, 2001). By these measures the reliability of the questionnaire was generally

poor, although between items, where a comparison could be made, there appeared to be better performance on the basis of the Cronbach's alpha values. However, Cronbach's alpha is strongly influenced by the number of items in the questionnaire, and using the approach of Treimen (2009) it can be shown that throughout the tables the two measures mentioned above are consistent for the number of items in the scales – 30 for the whole questionnaire and six for the factors. This is the case even though the datasets display negative skewness. So, accepting that the reliability was indeed poor, inspection of the results in greater depth was carried out and showed that the four reverse meaning questions had a major impact, yielding 39 negative correlations in the Pre-test and 33 in the Post-test, sometimes with quite large negative coefficients. One or two other questions also yielded several negative correlations. At this stage of the project there was clearly more work to be done in refining the questions to be used.

A feature of the results in Table 2.4 is the greater reliability of the questionnaire in the Post-test, relatively speaking, in particular on the PSD factors. There is an improvement of about 30% on the basis of median rho values, although this percentage varies considerably from factor to factor. The mean scores on questions did not change significantly from Pre-test to Post-test, therefore the pupils must have rearranged their scores (some up, some down) in order to arrive at the improved reliability. The two graphs in Figure 2.1 illustrate this phenomenon. An interesting question is whether this is a real signal or just a function of a poor questionnaire, and if it is a real signal, is it because the pupils became familiar with the questionnaire or because their visit to Benmore had a positive effect on their understanding of the issues addressed in the questionnaire.

Looking first at the possibility of questionnaire familiarity, the pupils' teachers and my research colleagues unanimously indicated that it is highly unlikely the pupils would be able to recall the contents of the questionnaire over an intervening period of two-weeks and certainly would not be aware that the questions grouped together into factors. This, then, would point to the possibility that pupils had a slightly better understanding in general of the relevance of the questions following their visit to Benmore Centre. Looking next at the possibility that the "improved" reliability is simply a function of a poor questionnaire, it will be instructive to inspect similar reliability measurements on subsequent Pre-test and Post-test administrations of the questionnaire in order to find out if improvement has been a consistent feature. In the following section the results of what was, in effect, a test-retest approach to reliability testing, provided by the Control Experiment, will be reported. Test-retest assumes

that the questionnaire will demonstrate equal reliability in both Pre- and Post-tests in the absence of an intervention. Intriguingly, this test-retest assumption implies that in the presence of an intervention a change of reliability should be observed, thus, we shouldn't have been surprised to see the change in reliability during the Pilot Study at Benmore that we have just briefly discussed.

It is noted that the item:total median rho coefficients for both the whole questionnaire and the individual factors are almost adequate. Since it is the total scores that will be used for analysis in the next Chapter, albeit for pupils not items, it is reassuring to see that these coefficients are a bit better than the item:item coefficients. These measures offer some indication of the robustness of the five factors built into the questionnaire. It was because the data from the experiment were slightly negatively skewed and the Likert scale was probably not interpreted by the pupils as a perfect interval scale (i.e. a score of 2 was not necessarily exactly two times a score of 1, etc) that an initial factor analysis was not carried out on the questionnaire. However, to try to understand better the performance of the questionnaire, a factor analysis using the SPSS principal component analysis facility with Varimax rotation was conducted at this stage. The results were mediocre at best, although better for the integrity of individual factors than for the questionnaire as a whole, and can be summarised as follows:

- the whole questionnaire only partially met the quality criteria for factor analysis, as recommended by Field, Chap.17 (Field, 2009), but all five factors met these quality criteria,
- the only intended factor consistently identified as a component in the Pre-test and Post-test analyses was Self-Perception of Competence (later in this thesis we will see that it is in this area of PSD that pupils feel most assured),
- Self-Perception of Competence and Awareness of Others were the strongest performing factors in both the Pre-test and the Post-test, in each case yielding only one component that included all five forward-meaning questions and excluded the one reverse-meaning question,
- Awareness of the Outdoor Environment consistently yielded two components, while Self-Perception of Confidence and Attitude towards Groups both improved from

yielding two components in the Pre-test to one component in the Post-test, provided reverse-meaning questions were ignored,

- apart from the instances just mentioned, there was only limited agreement as to which of the six items in a factor clustered on the components.

With regard to the quality of the intended factors, the fact that the mean and median correlation coefficients for the factors are only slightly higher than the equivalent values for the whole questionnaire is consistent with the factors barely emerging from the background level of correlations across the questionnaire. In other words, there were almost as many relatively good correlations outside factors as within them.

### **Further questionnaire development**

On the basis of the results from the initial validity, reliability and factor analyses just described, it was decided to replace all the reverse-meaning questions with forward meaning ones and carefully edit all the low-reliability questions. The revised version of the questionnaire was given to research colleagues for critical comment and then finalised with some minor adjustments.

The tight, two-year project schedule prevented any further revision of the questionnaire. Over autumn and winter 2009-2010 the Main Study data had to be collected. A Control Experiment was run in the winter to provide a test-retest of the final version of the questionnaire as well as a set of baseline results with which those from the Main Study could be compared. Because the experiment was run with the group of schools that participated in the Pilot Study, it is instructive to compare the performance of the revised questionnaire with the original as far as reliability is concerned.

The results of reliability tests from the Control Experiment are given in the same format as those from the Pilot Study (Table 2.5). The distribution of scores on items was negatively skewed to slightly greater extent than in the Pilot Study. Despite this limitation, Cronbach's alpha values are given here, as in the tables above.

The results in Table 2.5 bear some similarity to the results from the Pilot Study. Overall, using the median rho value as an indicator, the questionnaire performed about 25% better



than it did before, following the removal of the reverse-meaning questions. There were far fewer negative correlation coefficients in the coefficient matrix and the median rho values for item:total correlations are again approaching adequate. However, an improvement in reliability still exists between the Pre-test and the Post-test, even though this is a test-retest situation. The improvement is about 25%, and again it varies considerably from factor to factor: the Attitude towards Groups factor shows a marked improvement between tests in both the Pilot Study and the Control Experiment.

**Table 2.5** Measures of questionnaire reliability in the Control Experiment, for the whole questionnaire and within factors. Table above, Pre-test results; table below, Post-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$  is Cronbach's alpha, for comparison with the Spearman's rho values.

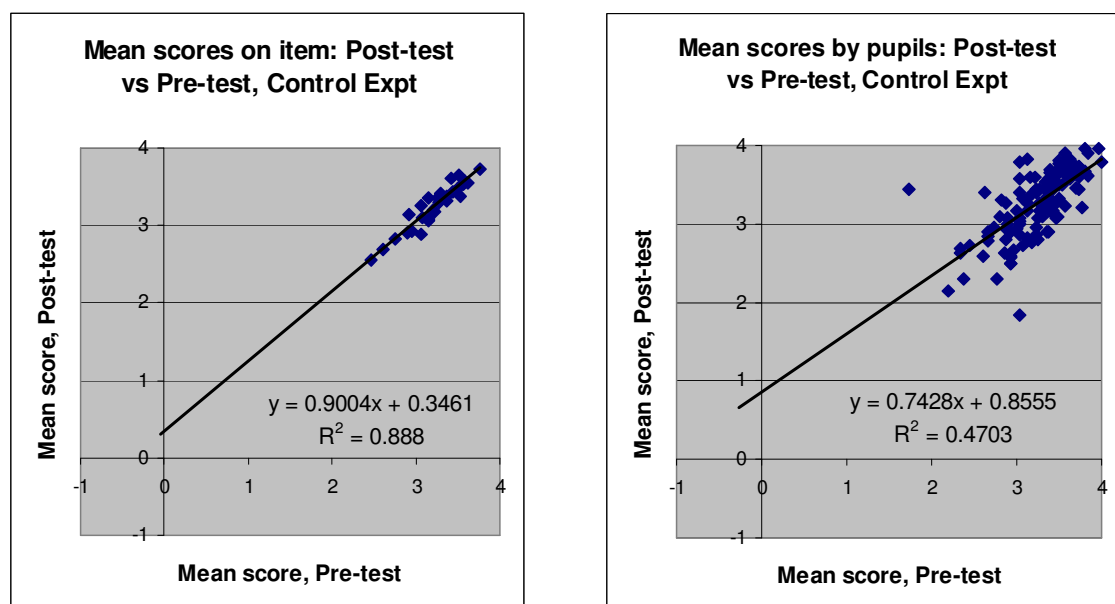
Items tested	Pre-test, item:item					Pre-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.18	0.19	0.71	50	0.89	0.43	0.45	0.61	97
Awareness of Outdoor Environment	0.31	0.28	0.71	86	0.72	0.63	0.68	0.74	100
Perception of Competence	0.26	0.24	0.53	67	0.71	0.58	0.57	0.68	100
Perception of Confidence	0.27	0.27	0.42	60	0.69	0.60	0.59	0.69	100
Awareness of Others	0.14	0.12	0.33	20	0.48	0.51	0.50	0.61	100
Attitude towards Groups	0.13	0.10	0.38	27	0.51	0.49	0.50	0.55	100

Items tested	Post-test, item:item					Post-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.24	0.25	0.75	72	0.91	0.50	0.50	0.67	100
Awareness of Outdoor Environment	0.30	0.29	0.75	73	0.74	0.63	0.70	0.74	100
Perception of Competence	0.24	0.23	0.45	67	0.66	0.59	0.58	0.66	100
Perception of Confidence	0.25	0.27	0.50	67	0.66	0.60	0.61	0.72	100
Awareness of Others	0.22	0.26	0.35	87	0.67	0.60	0.59	0.69	100
Attitude towards Groups	0.21	0.20	0.40	53	0.65	0.55	0.57	0.65	100

Given the consistency of performance of the questionnaire in the two experiments despite its relatively poor reliability, we can be more confident that there exists an improvement between Pre-test and Post-test that can be attributed to pupils' better understanding of the

questionnaire on their second response to it, especially in the areas of social development (“Others” and “Groups”). As in the Pilot Study, mean scores on questions did not change significantly from Pre-test to Post-test (Figure 2.2), therefore the pupils must have rearranged their scores (some up, some down) in order to arrive at an improved reliability. Given that in this case there was no intervention, it seems likely that the cause of the improvement is either questionnaire familiarity (despite what the teachers and researchers said) or a learning process that is not dependent upon there being an intervention. To investigate this interesting but unresolved issue a little further, the same tests of reliability were carried out on the data from the Main Study and are reported below.

**Figure 2.2** Left: strong linear relationship between mean scores on 30 questionnaire items, Control Experiment. Right: similar graph for the mean scores of 117 pupils.



In the test-retest situation provided by the Control Experiment a comparison between the Pre-test and the Post-test results is essentially a test of the external reliability of the questionnaire. A test of external reliability was carried out by using the split-half approach to reliability testing in SPSS and all 60 scores on items from the Pre-test and the Post-test, Pre-test scores in the first half and Post-test scores in the second half. The results are summarised, and compared with an equivalent analysis of the Pilot Study data, in Table 2.6a below. A non-parametric version of external reliability testing was attempted by calculating the Spearman’s rho coefficients for the correlations between factors and items within factors in the Pre-test and Post-test (Table 2.6b).

**Table 2.6a** External reliability coefficients from split-half analysis of the Control Experiment, Pre-test:Post-test, and a comparison with the same tests for the Pilot Study

	<b>Cronbach's alpha</b>	<b>Spearman-Brown Coeff, Equal Length</b>	<b>Guttman Split-Half Coeff.</b>
Control Expt	0.78	0.87	0.86
Pilot Study	0.59	0.74	0.74

**Table 2.6b** Spearman's rho coefficients for external reliability using the Control Experiment, Pre-test:Post-test.

<b>Factor</b>	<b>Correlation between factor, Pre-test : Post-test</b>	<b>Correlation between equivalent items within factor, Pre-test : Post-test</b>		
		<b>Mean rho</b>	<b>Median rho</b>	<b>Max. rho</b>
Awareness of Outdoor Environment	0.66	0.51	0.50	0.66
Perception of Competence	0.64	0.45	0.45	0.50
Perception of Confidence	0.67	0.44	0.39	0.62
Awareness of Others	0.58	0.50	0.54	0.63
Attitude towards Groups	0.41	0.41	0.40	0.54

The alpha values in Table 2.6a would be considered “adequate” by Garson (2009) for a split-half analysis, but we have seen that the data are not quite normally distributed, the histograms in Table 2.3 being typical of all datasets, and this is something of a limitation. The fact that the values for the Control Experiment are higher than those for the Pilot Study reflects stronger correlation between tests in the Control Experiment and less change in performance of the questionnaire from Pre-test to Post-test, as we would expect in a test-retest situation, notwithstanding the observations above on the internal reliability of each of the Pre-test and Post-test. The values in Table 2.6b for strength of correlation between the test and the re-test are better than those for correlations within any one test, but, overall, still fall short of an acceptable level of  $\geq 0.6$ . We can conclude from this external reliability study that the questionnaire is barely adequate.

A factor analysis of the data from the Control Experiment was also carried out. It used the SPSS principal component analysis facility with Varimax rotation. The results were poor, and can be summarised as follows:

- the analyses for both the whole questionnaire and the five factors met the quality criteria recommended by Field, Chap.17 (Field, 2009),
- using a factor loading of  $\geq 0.4$  as a threshold, the whole questionnaire yielded one component before rotation that included virtually all items, both in the Pre-test and the Post-test, but after rotation this broke down and it was not possible to identify any of the intended factors,
- apart from Self-Perception of Competence in the Pre-test and Awareness of Others in the Post-test, which yielded one component, all intended factors yielded two or more components after rotation,
- apart from the instances just mentioned, there was only limited agreement as to which of the six items in a factor clustered on the components.

In contrast to other tests of the quality of the questionnaire, this factor analysis suggests a slightly weaker performance in the Control Experiment compared with the Pilot Study. In any case, apart from some notable exceptions, such as the quality of the factor Self-Perception of Competence, there was considerable inconsistency as to which items clustered on to the components in both experiments. This is almost certainly due to the decision to construct the intended factors from questions that addressed a number of relevant traits rather than having every question addressing a single trait that would manifestly extract the same answer from pupils.

### **Questionnaire reliability in the Main Study**

Here the same reliability test results are shown for the Pre-, Post- and Delayed-test of the Main Study, using about 350 pupils (Table 2.7). The Pre- and Post-tests were conducted the week before and the week after the residential week of OAA and the Delayed-test 2-3 months later.

**Table 2.7** Measures of questionnaire reliability in the Main Study, for the whole questionnaire and within factors. Table above, Pre-test results; middle table, Post-test results; table below, Delayed-test results. % sig. is the percentage of correlations that have statistical significance; C $\alpha$  is Cronbach's alpha, for comparison with the Spearman's rho values.

Items tested Pre-test	Pre-test, item:item					Pre-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.19	0.19	0.68	84	0.88	0.44	0.45	0.56	100
Awareness of Outdoor Environment	0.29	0.23	0.68	67	0.73	0.61	0.70	0.74	100
Perception of Competence	0.20	0.19	0.27	20	0.62	0.54	0.53	0.60	100
Perception of Confidence	0.24	0.22	0.33	47	0.64	0.59	0.60	0.66	100
Awareness of Others	0.25	0.24	0.32	46	0.68	0.60	0.61	0.65	100
Attitude towards Groups	0.16	0.17	0/39	20	0.56	0.52	0.52	0.60	100

Items tested Post-tset	Post-test, item:item					Post-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.24	0.25	0.61	93	0.91	0.49	0.51	0.62	100
Awareness of Outdoor Environment	0.35	0.36	0.61	80	0.77	0.64	0.72	0.77	100
Perception of Competence	0.26	0.26	0.42	59	0.71	0.59	0.59	0.66	100
Perception of Confidence	0.32	0.32	0.45	67	0.73	0.63	0.67	0.70	100
Awareness of Others	0.29	0.30	0.37	46	0.73	0.62	0.63	0.67	100
Attitude towards Groups	0.27	0.28	0.42	20	0.70	0.59	0.60	0.68	100

Items tested Delayed-test	Delayed-test, item:item					Delayed-test, item:total			
	Mean rho	Median rho	Max. rho	% sig.	C $\alpha$	Mean rho	Median rho	Max. rho	% sig.
Whole q'naire	0.26	0.27	0.65	95	0.92	0.52	0.53	0.66	100
Awareness of Outdoor Environment	0.38	0.30	0.65	100	0.79	0.67	0.62	0.81	100
Perception of Competence	0.32	0.32	0.42	100	0.75	0.63	0.64	0.67	100
Perception of Confidence	0.27	0.27	0.38	100	0.71	0.61	0.61	0.65	100
Awareness of Others	0.30	0.31	0.42	100	0.74	0.64	0.64	0.64	100
Attitude towards Groups	0.30	0.29	0.48	93	0.74	0.61	0.63	0.61	100

**Table 2.8** Descriptive statistics for pupils' mean scores in the Main Study.

	<b>Pre-test mean</b>	<b>Pre-test standard deviation</b>	<b>Post-test mean</b>	<b>Post-test standard deviation</b>	<b>Delayed- test mean</b>	<b>Delayed-test standard deviation</b>
Boys	3.32	0.38	3.38	0.42	3.24	0.47
Girls	3.44	0.33	3.51	0.36	3.39	0.38

In the context of the results and discussion for the Pilot Study and the Control Experiment, these results are interesting. An important point to make to begin with is that the descriptive statistics for this study are different (Table 2.8). The mean score for both boys and girls are higher than in the Pilot Study and Control Experiment, and at the same time, as will be shown in Chapter 3, there is a small but statistically significant increase in scores from Pre-test to Post-test followed by a moderate and statistically significant fall to the Delayed-test. These changes in mean score are accompanied by an increase in median rho values between Pre-test and Post-test of about the same magnitude as in the Pilot Study and Control Experiment, and thus establish a typical behaviour of the questionnaire as far as its reliability is concerned. This behaviour seems to be independent of whether there is an intervention and of the values of mean scores. The results for the Delayed-test, taken 2-3 months later, show very little difference in reliability from the Post-test, confirming that a significant change in the values of mean scores (a fall in this case) has no effect. In the discussion above we questioned whether the teachers' and researchers' view, that pupils would not remember the contents of the questionnaire over a two-week period, is correct, thus permitting a measure of questionnaire familiarity to influence reliability between administrations. The information that we now have, that there was little or no change in reliability over a 2-3 month period, would be consistent with a measure of familiarity being retained over a few weeks but not over a few months. This is not to say that the pupils remembered individual questions – the teachers' and researchers' comments will be correct in this respect – but it is possible that the questionnaire acted as an *aide memoire* to the revision of what the pupils already know about the outdoor environment and PSD, leading to slightly more convergent answers when tested just a couple of weeks later. This interpretation will be revisited in Chapter 4, "Implications for Research Methodology"

As with the other experiments, despite limitations on its significance, a factor analysis was carried out on the three tests in the Main Study. On this occasion the results were somewhat better than in the previous experiments, not least because the sample size is now about 350, rather than c.100, and all quality criteria were firmly met. As we have seen previously, for the whole questionnaire a single component is identifiable before Varimax rotation, but this

breaks down after rotation and it is difficult to identify the intended factors. For the five intended factors, they perform quite strongly, yielding a single component in all three administrations of the questionnaire, with the exception of two components for “Attitude towards Groups” in the Pre-test and “Awareness of the Outdoor Environment” in the Delayed-test.

It should be pointed out at this stage that some work was carried out to discover if any subsets of the six questions within each intended factor would perform better in terms of the Spearman’s rho coefficients they generated. The factor analyses were used to help in the identification of likely subsets. Some improvement was seen in the Pilot and Control experiments, but for the Main Study, where the factors performed relatively well, little improvement was seen. On this basis, analysis of the Main Study data was pursued using the intended six-item factors.

## **2.4 Summary of Chapter**

Rather more time was spent on questionnaire design in this project than was intended. The initial product, containing five intended factors that would address the aims of the project, were made up of rather wide-ranging questions instead of the very focused ones typical of most questionnaires used in quantitative studies, such as the LEQ of Neill (2007). This questionnaire performed poorly in terms of reliability tests with the chosen measure of performance (Garson, 2009), but in the process of testing, some intriguing results emerged, and thus the questionnaire itself became a research project. Unfortunately, the time constraints of the overall project demanded that the main data set be collected. Therefore, the Main Study, on which the data analysis in Chapter 3 is based, was carried out using the questionnaire with no further improvements.

The account of experimental design explains the sequence of questionnaire administrations and the time constraints on these. It also explains how two groups of schools were used, of necessity, but this also led to a research outcome that had not been planned, concerning the effects of socio-economic background on the scores of pupils, which is fully described in Chapter 3.

The reliability tests on the questionnaire revealed an improvement in performance between the Pilot Study and the Control Experiment of about 20-30%, which was expected since

there were changes to several questions on the basis of the Pilot Study. However, within each of these experiments the reliability also increased by 20-30% between the Pre-test and the Post-test, raising the possibility of some kind of questionnaire familiarity, contrary to the views of the pupils' teachers. The Control Experiment, with no intervention, and the Main Study, with an intervention, were carried out with the same questionnaire but different groups of schools. The reliability tests for the Main Study showed a performance similar to that in the Control Experiment, thus suggesting that any questionnaire familiarity would not be the result of an intervention (unless an intervention effect and a school group effect cancelled each other out). The Delayed-test of the Main Study, 2-3 months later, more-or-less maintained the performance level for reliability at the Post-test level. This evidence, taken together with the teachers' view that pupils would not remember the contents of the questionnaire even over a two-week period, suggests that any familiarity effect is not at the level of individual questions but at the level of general concepts. These PSD and environmental concepts are frequently-visited components of the school curriculum, and it would seem entirely reasonable that exposure to the questionnaire acted as an *aide memoire* of these concept, such that scores on questions in second and later administrations of the questionnaire converged better within factors.



## **Chapter 3            Data Analysis**

### **3.1      Introduction**

At the level of the peer-reviewed publication in outdoor education it is rare to see an in-depth analysis of project data and the statistical tests used on it to arrive at meaningful results that can be confidently interpreted. This is not surprising, given the premium on space in academic journals. Summary results are often published, however, and it is often necessary to take it on trust that these have come from rigorous analyses of the data. In contrast, a research thesis like this offers an opportunity to report on extensive, in-depth analyses of data, statistical tests and modelling attempts. Readers can confidently judge whether the results, interpretations and conclusions are valid.

The principal data set reported here is made up of pupils' scores on the three administrations of the questionnaire in the Main Study: the Pre-test, conducted in the week before the pupils left for their residential week of OAA, the Post-test, conducted in the week after they came back, and the Delayed-test, conducted 2-3 months later. For purposes of comparison with the results from the Main Study data, the scores on the same questionnaire from the Control Experiment, in which there was no intervention, are also reported here. This experiment had only a Pre-test and Post-test, two weeks apart. Details of data collection are given in Section 2.2.2. The questionnaire comprised 30 questions (items) grouped into five factors and was described in Section 2.3.

Initially, a thorough inspection of the data, assembled in various forms, is presented in order to identify features of the data set, such as trends, differences or common occurrences in the data that might be relevant to the objectives of this project and amenable to statistical tests. One or two statistical descriptives are used to help the inspection, but because the data are negatively-skewed in their distribution there is a limit to the use of the usual parametric descriptives. The inspection is carried out first on scores averaged for items and then on scores averaged for pupils. The average, or mean, score is felt to be a meaningful descriptive because even with skewed data it gives a summary indication of the relative strength of scores from pupil to pupil and test to test.

Following the description of the mean scores and the tests for the statistical significance of differences in means, a comprehensive linear regression analysis of data is presented. The

regressions are to investigate cause and effect and the power of prior PSD skills amongst pupils to predict scores following the OAA intervention. It is wise to be wary of regression analysis on skewed data, given that the tail to the data is only one-sided, but provided there is a normal distribution to the residuals (residual displacements) of scores from the least-squares best-fit straight line, then, technically, the regression is likely to be acceptable. SPSS also provides a facility for testing whether a few scores in the tail of the data (in this case towards low scores) have an undue influence on the regression.

At the end of each section of this chapter there are a few sentences on the possible relevance of the results for the aims and objectives of this project. However, in view of the modest quality of the questionnaire as an instrument for testing self-perception of PSD skills, and some inherent limitations in the survey method, all observations are tentative. Nevertheless, they will be carried forward to later, discussion chapters (Chapters 4, 5 and 6) and interpreted in wider contexts.

### **3.2 Data management and software**

Questionnaires were completed by pupils by simply ticking their responses to 30 questions (in reality, statements about themselves) on a 4-point Likert scale. Apart from the two administrations of the questionnaire for the Pilot Study, there were 3 administrations for the Main Study and two for the Control Experiment. Scores from 1 to 4 were transferred manually to Microsoft Excel spreadsheets, and from there imported into the PASW Statistics 17 module of SPSS (Statistical Package for the Social Sciences). These software packages are well known to social scientists and will not be described here.

Some basic data management and statistical work was carried out in Excel, and some graphs produced in Excel. For example, the division of scores into tenths during a comparison of Pre-, Post- and Delayed-test scores in the Main Study (see Section 3.4.1) was carried out in Excel. However, the vast majority of statistical work was carried out in SPSS. The textbook by Field (2009) has been invaluable as a guide to understanding and working with SPSS.

### **3.3 Data inspection and descriptives**

Data inspection is simply an inspection of the scores by pupils on items in the questionnaire at various stages of the project (in this case, Control Experiment and Main Study, but not the

Pilot Study) to identify trends that might be meaningful for the aims of this research and, thus, worthy of further study for effect size and statistical significance. Most of the inspection is carried out on mean scores, either mean scores on items or pupils' mean scores. As such, this addresses the descriptives of the data, but as will become evident, the scores have non-normal (negatively skewed) distribution and this restricts the use of parametric descriptives to summarise the scores. The mean is used here since its statistical significance is well known and it provides an immediately understandable indication of shifts in scores from, for example, the Pre-test to the Post-test administration of the questionnaire. An exception to using only the mean occurs in the next section, where a brief inspection of the distribution of scores on items is presented and the mode proves to be a useful descriptive.

Two general points should be made here. First, in addition to considering datasets at the level of the whole questionnaire (30 items) and at the level of the intended factors (6 items in each of five factors), it is appropriate to create a dataset that includes only the PSD-related items (24 items). This is important because it is the measurement of the benefit that is reflected in the change in scores on the PSD items that sits at the heart of this research. Moreover, it is intended to investigate the possibility that pupils' awareness of the outdoor environment influences their potential to benefit from a residential week of OAA. Therefore, both the 24-item PSD score and the 6-item "Awareness of the Outdoor Environment" score appear throughout the following discussion. Second, it is possible to make a lot of detailed, in-depth observations with inspections of this kind, and over-interpret the data. An effort is made to be sensible, and restrict observations to the broadest features of the data.

### **3.3.1 Distribution and significance of scores on items**

This inspection is carried out for the Main Study and the Control Experiment. It illustrates the non-normal distribution of scores and highlights broad differences in responses to questions across Pre-, Post- and Delayed-tests and within intended factors.

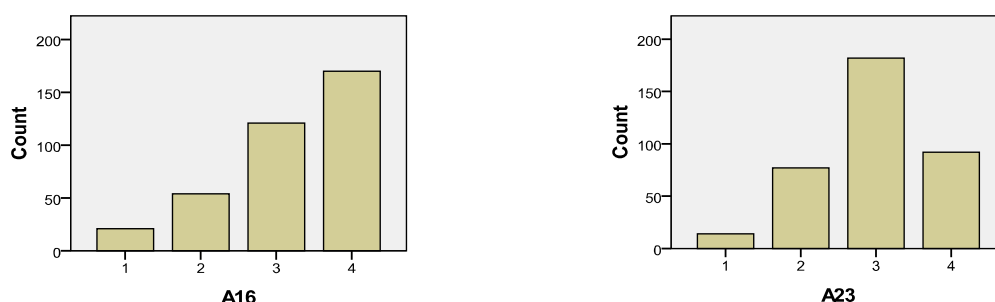
#### **Distribution of scores on items**

The distribution of scores on the 1-4 Likert scale is presented as histograms in Appendix 3.1. The mean score on each question is also given. It is important to draw a distinction between these scores, aggregated up for items, and the scores aggregated up for pupils (cases), which

are described in the next section and are the relevant measure for assessing questionnaire reliability and understanding personal and social development. Nevertheless, as a first inspection of these scores, item modes and means highlight some interesting features of the data, as follows.

- Almost without exception the scores are towards the high end of the Likert scale and their distribution negatively skewed towards low values. This is also a feature of the data aggregated up for pupils. Analysis and modelling of the data for pupils is potentially restricted because of this, as will become apparent in due course (principally, in Section 3.4.1).
- The distribution of scores on any one question is broadly the same in all three of the Main Study tests and in the two tests of the Control Experiment. There are some departures from this, but not in any systematic way to suggest further investigation at this stage. It is reassuring that the questionnaire yields these similar patterns under five different administrations, with the *proviso* that we cannot be sure from these scores on items that individual pupils are indeed scoring so consistently. At the same time, but less encouraging, the striking similarity in distribution suggests a lack of intervention effects in the data.
- If we adopt as a model a pattern of increasing score count from 1 to 4 as a reference (e.g. question A16 in Figure 3.1), then some noteworthy exceptions to this can be identified.

**Figure 3.1** Reference distributions of scores on items. Left: the count of scores increases progressively from 1 to 4 on the Likert scale. Right: 3 is the modal score.



First, Question 6 is unique, and clearly shows that only c.50% of pupils find nature programmes on TV interesting. It is interesting to note, however, that despite its unique distribution of scores, question 6 factors in with the set of six questions on the outdoor

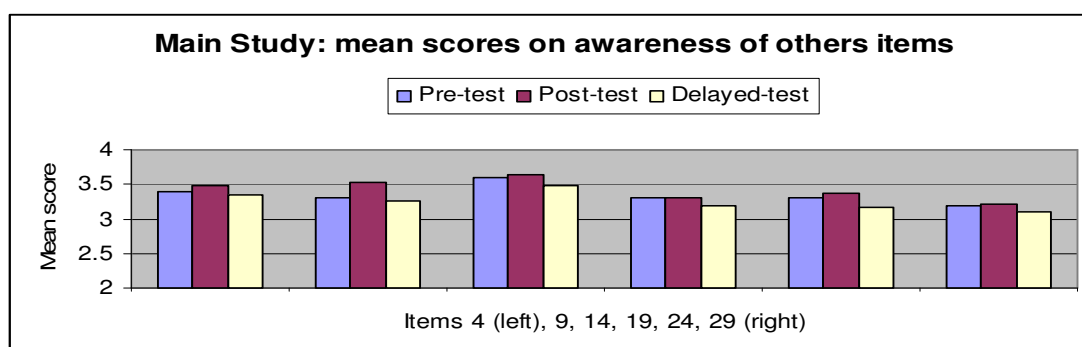
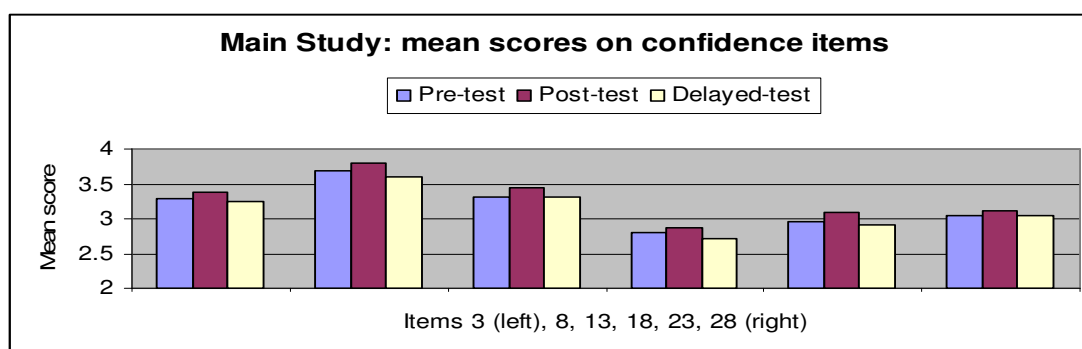
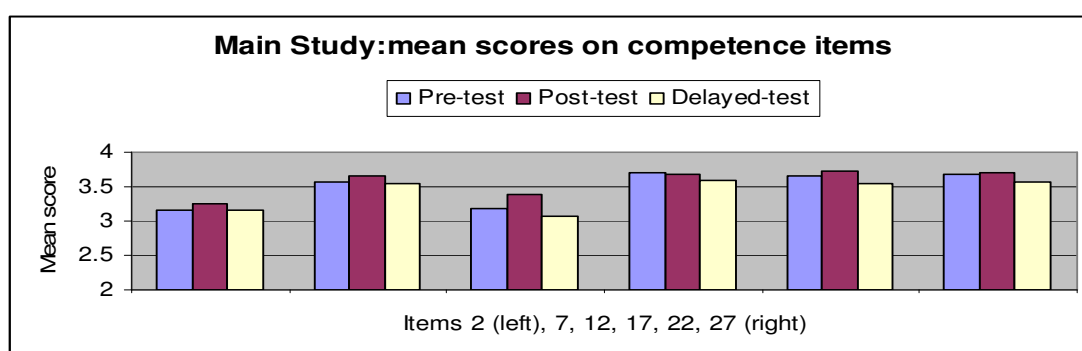
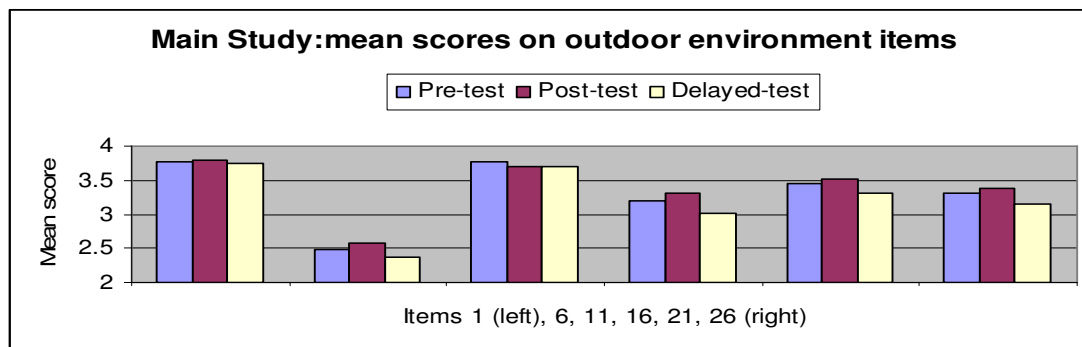
environment. This must be because pupils who score low on question 6 also score low on the other questions within this factor, and *vice versa*.

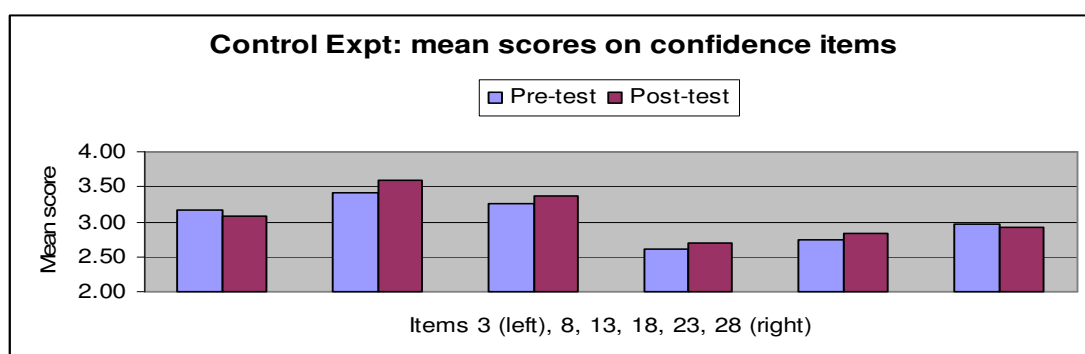
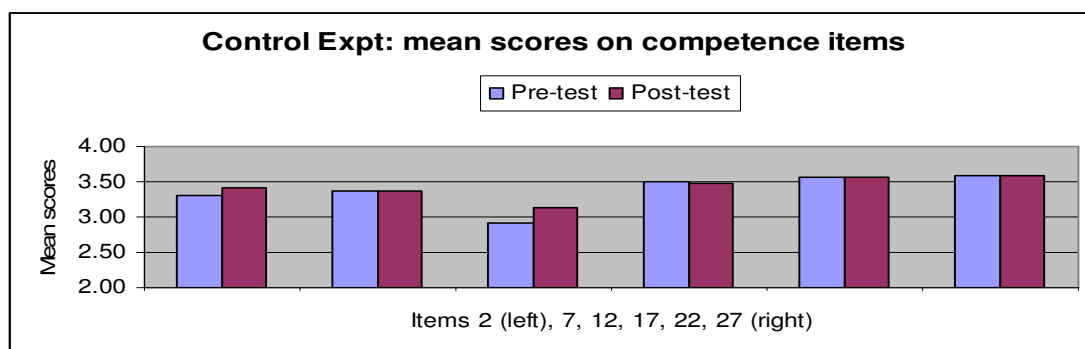
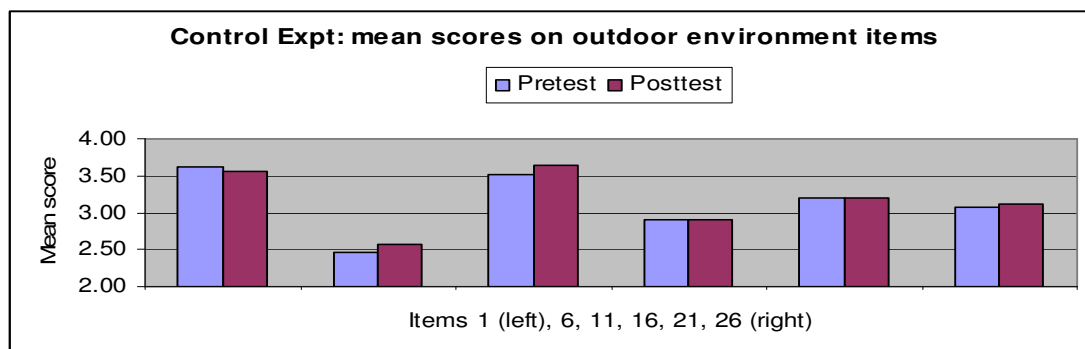
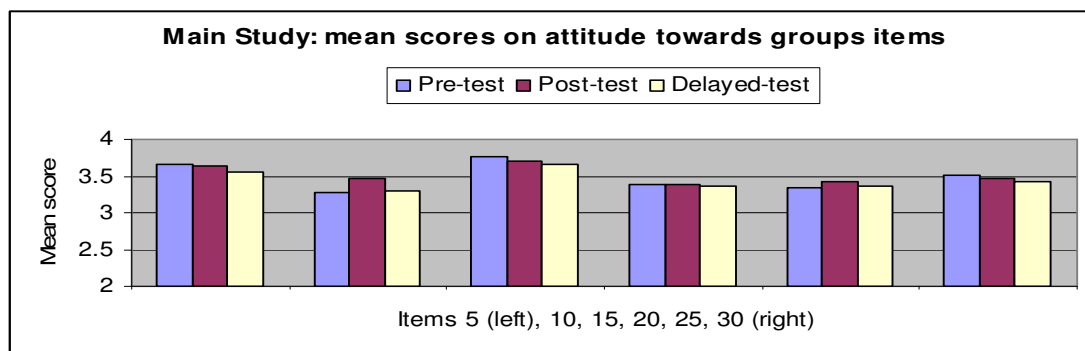
Second, there are several questions on which a score of 3, rather than 4, is the mode (e.g. question A23 in Figure 3.1). These are often the same questions in the Main Study and the Control Experiment, e.g. Q10, Q23. The vast majority of these questions occur within the factors addressing pupil confidence and their attitude towards others, where mean scores are correspondingly lower. In the Main Study the majority are in the Delayed-test, where pupils have scored lower than in either the Pre-test or the Post-test. Indeed, the Post-test is almost devoid of questions that have 3 as their mode. By contrast, in the Control Experiment they are approximately equally divided between Pre- and Post-tests.

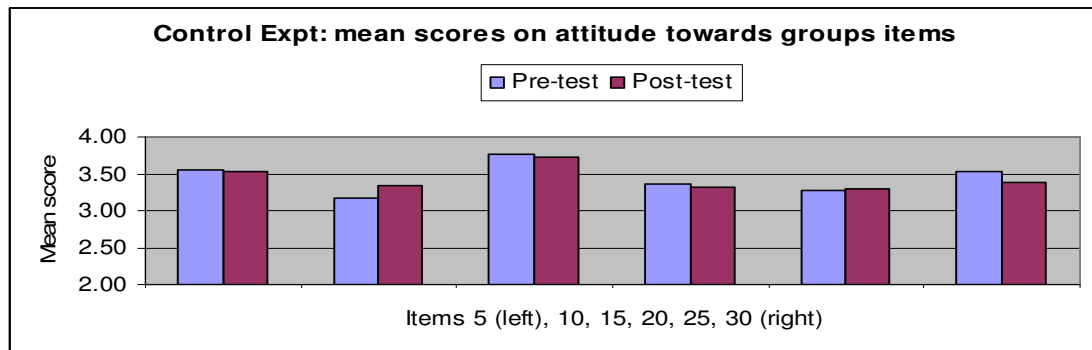
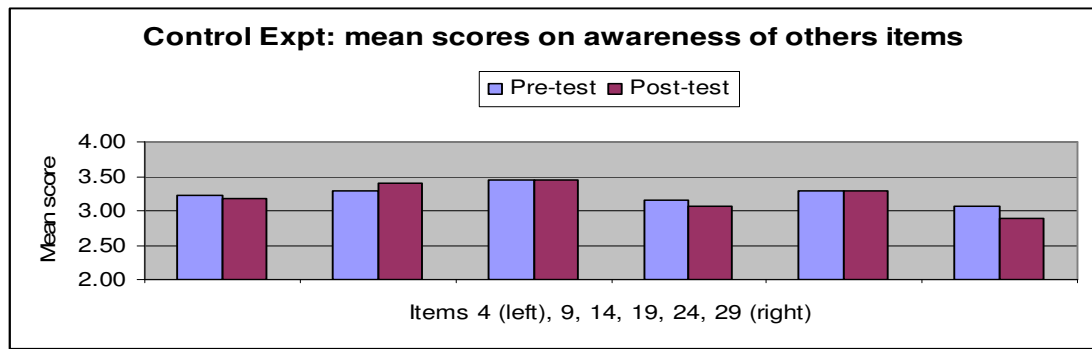
Turning now from the distribution of scores to the mean scores on items, some clear trends can be identified. In addition to the mean values attached to the histograms, Figure 3.2 shows how mean score for each question changes from Pre-test to Post-test to Delayed-test in the Main Study and from Pre-test to Post-test in the Control Experiment. We can make the following observations.

- In the Main Study, means tend to increase from the Pre-test to the Post-test, but then decrease markedly in the Delayed-test to a level that is commonly lower than the Pre-test. This is clearest on the items addressing “Perception of Confidence”. Focusing on the four factors addressing PSD, if this trend is confirmed by the pupils’ mean scores it raises two possible interpretations: an improvement in perceived PSD attributes from before to after the OAA week, but a complete loss of this improvement three months later; and/or a residential week effect in which the scores immediately prior to the week and, in particular, immediately after, are inflated by excitement and a sense of wellbeing amongst pupils relative to a norm as illustrated by the Delayed-test.
- In contrast to the Main Study, the Control Experiment yields more-or-less equal numbers of questions for which the mean score decreases in the Post-test as increases, and many questions on which there is no change. Optimistically, we might interpret this to demonstrate rather random shifts of scores associated with there having been no intervention.

**Figure 3.2** Column charts of changes in mean scores on items across five applications of the questionnaire in the Main Study and the Control Experiment, grouped by intended factor. Note that the two experiments were conducted with pupils from two different groups of schools.







- The mean scores on items in both the Main Study and Control Experiment show the same highs and lows, e.g. Q1 has a high mean in both and Q18 has a low mean in both. It is reassuring to see similar responses to questions across the two groups of schools involved in the study. However, despite similar relative scores, the means are lower in the Control Experiment by an average of 0.15 (c.4.5%). A Wilcoxon Rank-Sum Test of this difference is reported in the next section.

### Statistical significance of scores on items

Following these observations, Wilcoxon Signed Rank Tests for non-parametric, related samples were carried out on the item means from the Main Study and the Control Experiment to discover whether the differences in values outlined above are statistically significant. Effect size,  $r$ , is calculated from the  $z$  score using the equation given in Field (2009):

$$r = z / \sqrt{N}, \quad \text{where } N \text{ is the total number of observations.}$$

Field uses Pearson's correlation coefficient for effect size throughout his book, for both parametric and non-parametric statistics. In the tables below, statistically significant small (0.2) to medium (0.5) effect sizes are shown in bold type.



**Table 3.1** Results and outcomes from statistical tests on mean scores on items for both Main Study and Control Experiment.

Test		Effect size, $r^*$	2-tailed significance, $p$	Outcome
Main Study, between Pre-test and Post-test	All 30 items	<b>0.50</b>	0.000	Post-test item means are greater than Pre-test item means by a statistically significant difference of medium to large size on PSD factors.
	“Awareness of Outdoor Environment” only (6 items)	0.52	0.074	
	PSD items only (24 items)	<b>0.51</b>	0.000	
Main Study, between Pre-test and Delayed-test	All 30 items	<b>0.57</b>	0.000	Delayed-test item means are smaller than Pre-test item means by a statistically significant difference of medium size.
	“Awareness of Outdoor Environment” only (6 items)	<b>0.64</b>	0.028	
	PSD items only (24 items)	<b>0.55</b>	0.000	
Main Study, between Post-test and Delayed-test	All 30 items	<b>0.61</b>	0.000	Delayed-test item means are smaller than Post-test item means by a statistically significant difference of medium size.
	“Awareness of Outdoor Environment” only (6 items)	<b>0.59</b>	0.042	
	PSD items only (24 items)	<b>0.62</b>	0.000	
Control Expt, between Pre-test and Post-test	All 30 items	0.09	0.487	There is no statistically significant difference between the Pre-test and Post-test item means.
	“Awareness of Outdoor Environment” only (6 items)	0.27	0.345	
	PSD items only (24 items)	0.05	0.715	

\* According to Cohen (1988, p.26), a medium effect size of 0.5 is “large enough to be visible to the naked eye”!

A Wilcoxon Rank-Sum Test was carried out on the two independent samples of 30 items each ( $N=60$ ) from the two different school groups in the Main Study and the Control Experiment to find out if a statistically significant difference exists in their item means that could be attributable to a school group effect. A school effect could be socio-economic in origin, related to the ethos of the schools or simply related to the excitement associated with the adventure week experienced by the pupils in the Main Study. The Test was carried out on the data of the two Pre-tests (pre-intervention in the Main Study). The average of 0.15 by which the item means in the Control Experiment are smaller than those in the Main Study yielded a small to medium effect size of  $r=0.23$ , with a 2-tailed significance of  $p=0.078$ ; in other words, the small effect is not significant. The possibility of a school effect in the data will be re-examined using pupil means.

### 3.3.2 Distribution and significance of scores by pupils

These scores are the ones that will be used for hypothesis testing, modelling and interpretation in terms of personal and social development at the late primary stage of education. In particular, the changes in scores from Pre-test to Post-test to Delayed-test in the Main Study, where there has been an intervention, and how these compare with scores in the Control Experiment, where there has been no intervention, are important, with the *proviso* that two different groups of schools were used and a school effect might have to be taken into account. Self perception of PSD is given by the mean scores for the 24 PSD items, which are grouped into four factors, and Awareness of the Outdoor Environment is given by the mean scores for the 6 items in that factor.

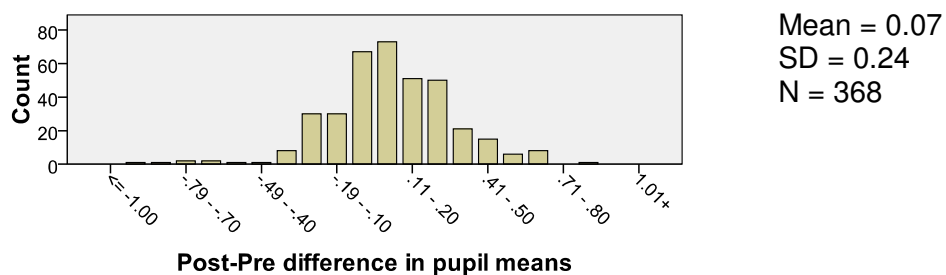
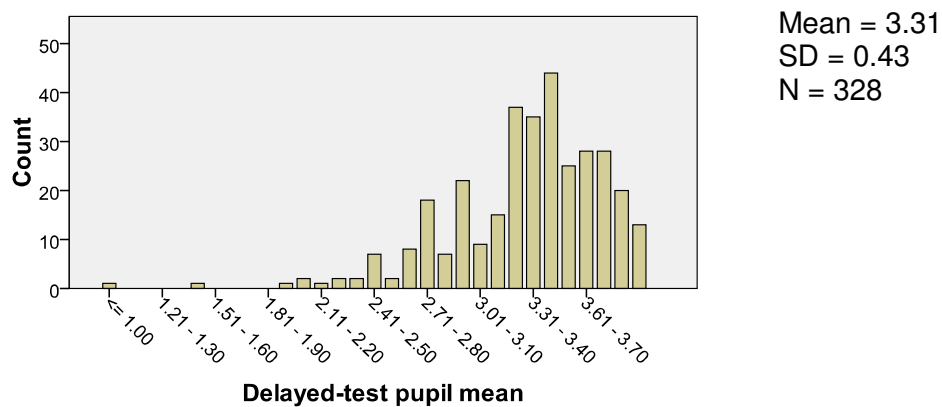
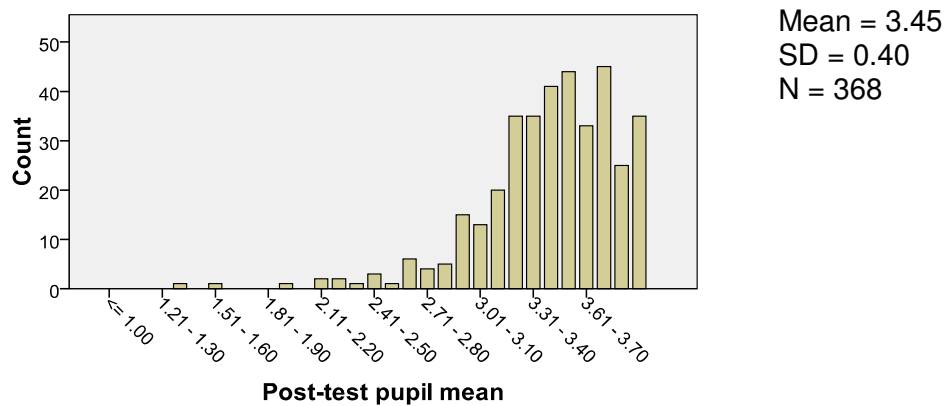
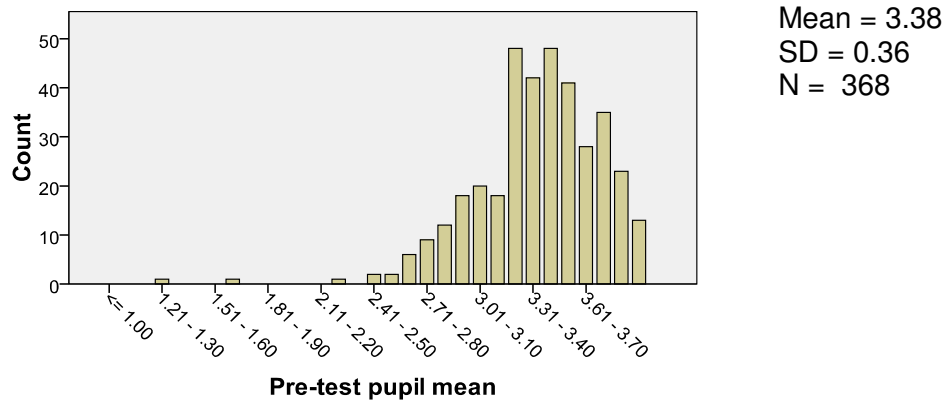
#### Pupils' scores on personal and social development items

Figure 3.3 shows histograms of the distribution of pupils' mean PSD scores for all the tests in the Main Study and Control Experiment, in total and for boys and girls. In Appendix 3.2, the same results are presented for each factor. Some general observations and inferences can be made from these distributions of pupil means, as follows.

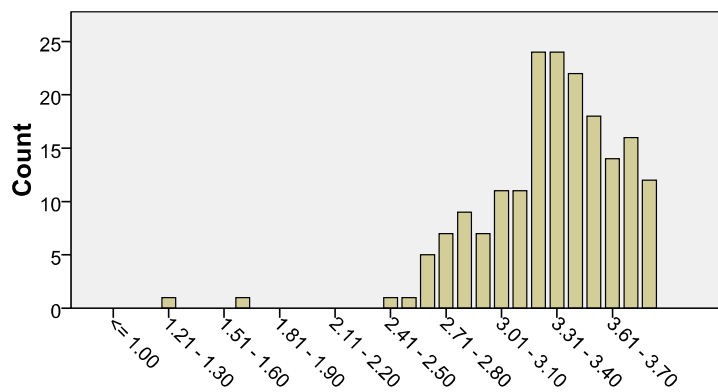
- The highly skewed nature of the data is confirmed. Differences in scores between tests, e.g. between the Pre-test and Post-test of the Main Study, shown in Figure 3.3 as an example, are distributed normally, however. This might be helpful at a later stage for some statistical tests.
- The negative skewness in all the data sets is caused chiefly by low scores amongst boys. In fact, some of the data sets for girls test normal using the Kolmogorov-Smirnov and Shapiro-Wilk tests, namely the Pre-test and Post-test distributions in the Control Experiment and, less convincingly, the Pre-test distribution in the Main Study. It is clear that some boys are not at all confident of their personal and social skills.
- The means for all PSD items for the whole sample of the Main Study confirm the increase in scores from Pre-test to Post-test (1.8%) and then a strong decrease to the Delayed-test (-3.3%). In contrast, the difference in scores between Pre-test and Post-test in the Control Experiment is <1%. Some comments on how this might be interpreted are given above in the section on scores on items.

**Figure 3.3** Distribution of pupils' mean scores on PSD items for the Main Study and Control Experiment, for all cases and for boys and girls. SD = standard deviation; N = size of sample. N.B., the Count (vertical) scale changes with sample size.

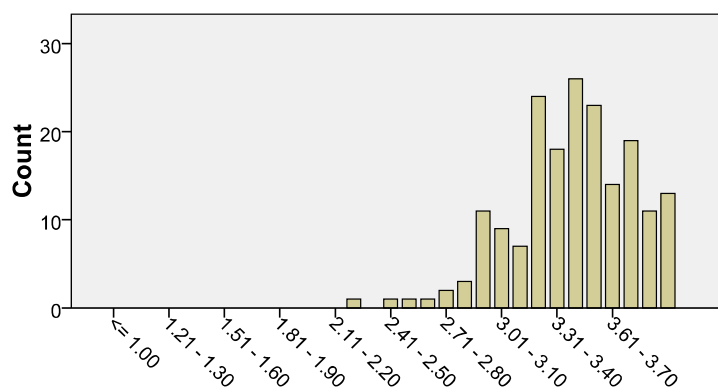
**Main Study, all cases**



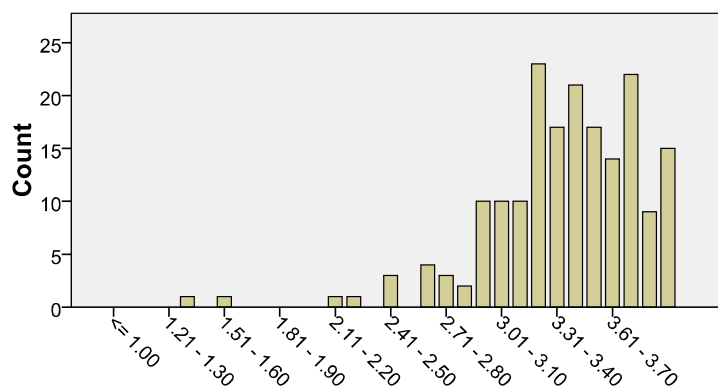
## Main Study, by gender



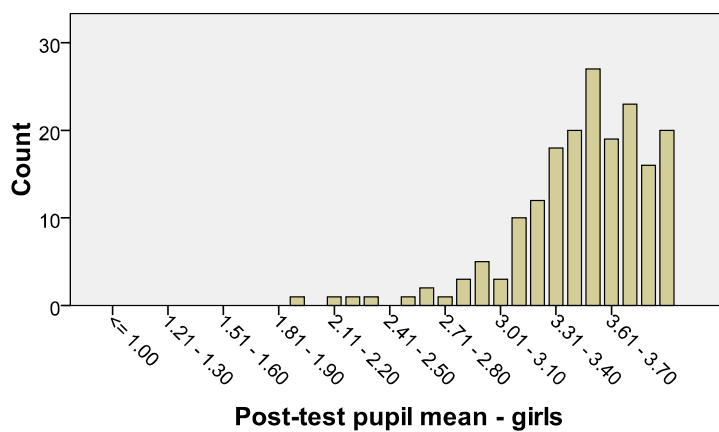
Mean = 3.32  
SD = 0.38  
N = 184



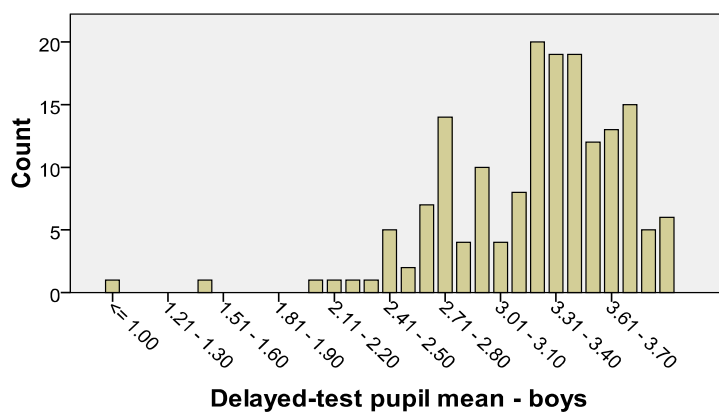
Mean = 3.44  
SD = 0.33  
N = 184



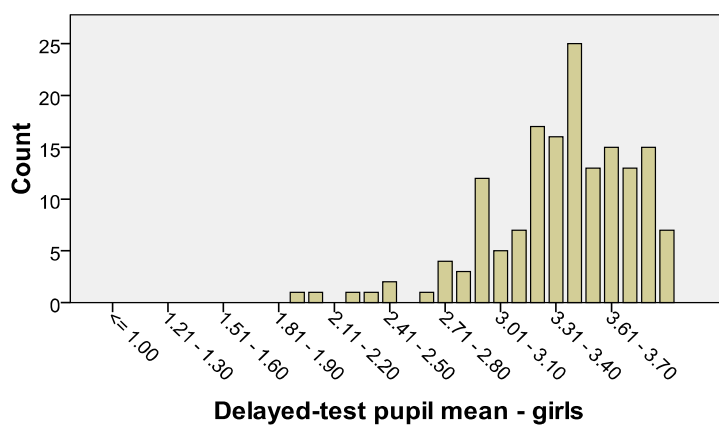
Mean = 3.38  
SD = 0.42  
N = 184



Mean = 3.51  
SD = 0.36  
N = 184

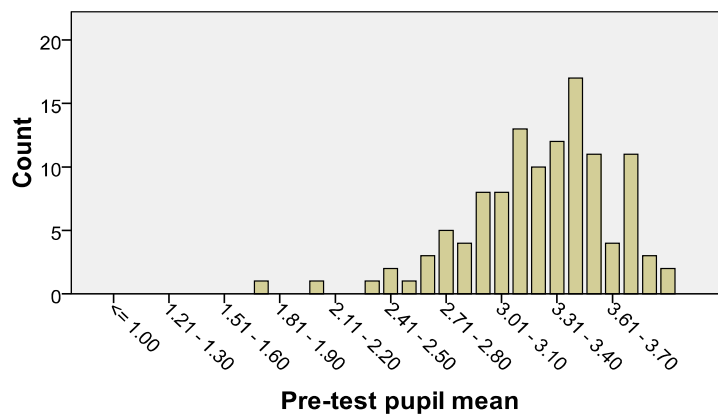


Mean = 3.24  
SD = 0.47  
N = 169

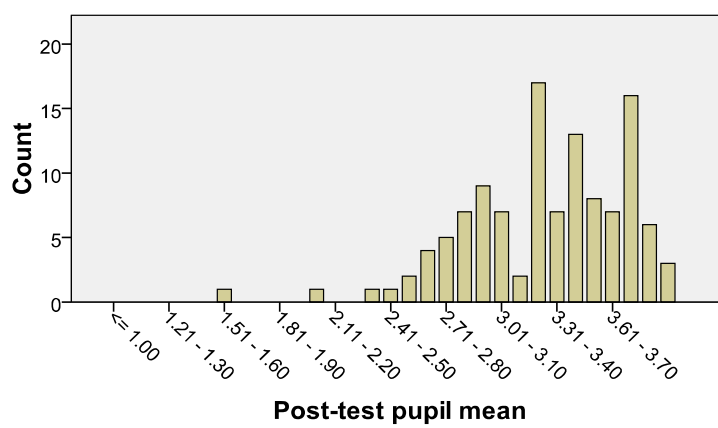


Mean = 3.39  
SD = 0.38  
N = 159

### Control Experiment, all cases

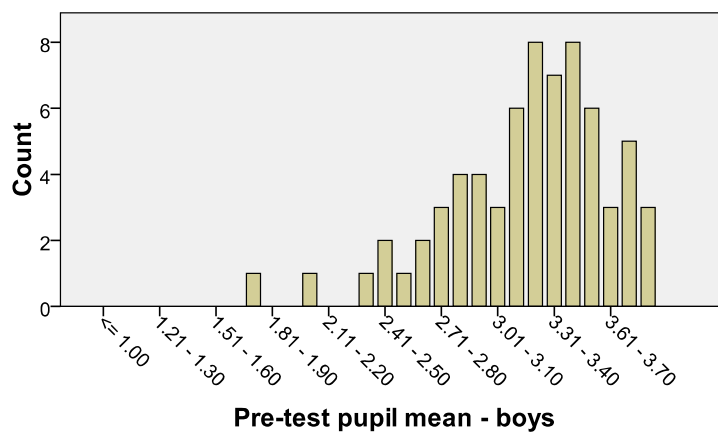


Mean = 3.26  
SD = 0.39  
N = 117

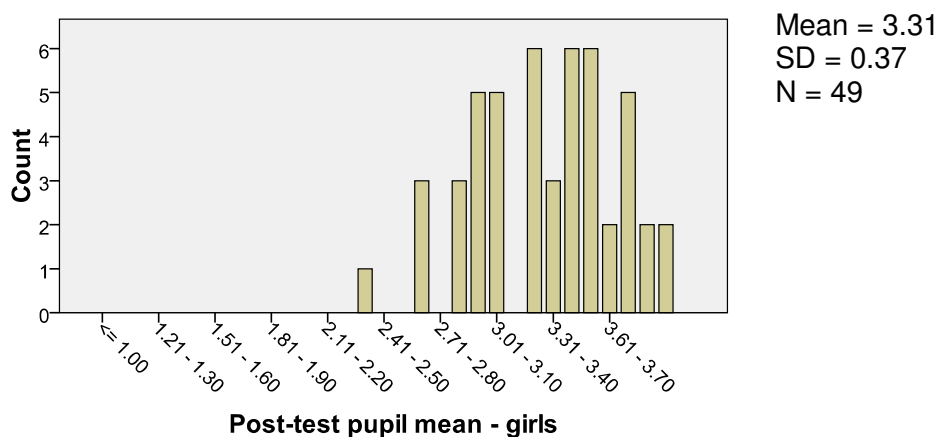
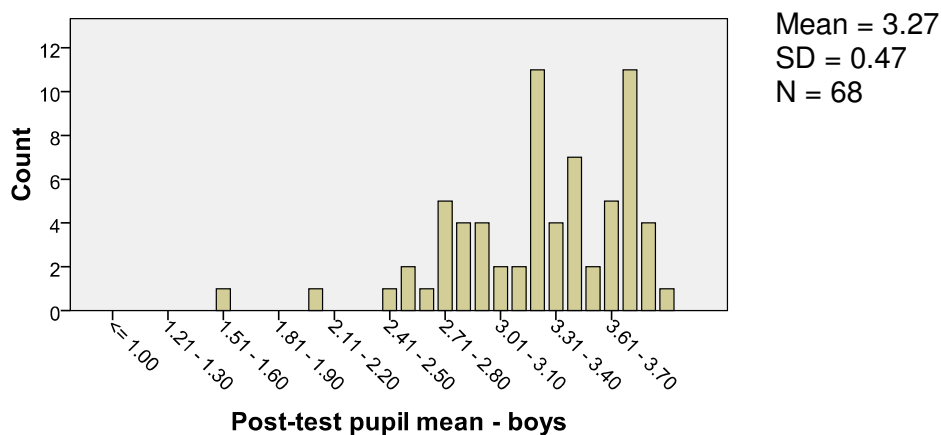
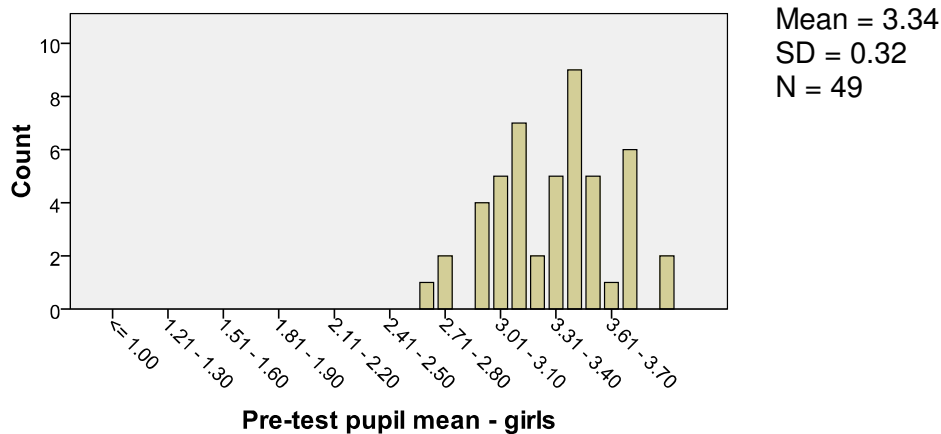


Mean = 3.29  
SD = 0.43  
N = 117

### Control Experiment, by gender



Mean = 3.21  
SD = 0.43  
N = 68



- Now that we are inspecting the pupils' mean scores, gender differences are apparent. In the Main Study, both boys and girls scores follow the same trend from Pre-test to Post-test to Delayed-test (an increase and then marked decrease), but, on PSD items, girls score higher than boys by about 0.13 (c.4%) on average. The fact that this difference in scores between genders across the three tests stays very much the same suggests that the phenomenon is real for this group of pupils, and that intervention does not affect gender differences. Even at the level of individual factors the essential features of this pattern

still exist. It might be surmised from this situation that girls have a more positive view of their personal and social skills at the late primary school stage.

- This picture is not quite as clear in the Control Experiment, however, where the girls score drops from the Pre-test to the Post-test, all scores being lower than in the Main Study by between 3% and 6%, as noted under consideration of items. At the level of factors (Appendix 3.2), the picture is confirmed, with girls more often than not scoring lower in the Post-test than in the Pre-test. The combination of lower scores throughout and less positive feelings amongst girls again raises the possibility that a school effect is influencing the outcome. Alternatively, the inconsistent pattern of scores in the Control Experiment might reflect the fact that pupils were not anticipating or benefiting from a residential week of activities at an outdoor centre.
- Broken down by factor in Appendix 3.2, it becomes clear that the distribution of mean scores by pupils in Perception of Confidence is different from that in other factors. Mean scores show that confidence is relatively low in all administrations of the questionnaire in both the Main Study and Control Experiment, and this is the case for both boys and girls. Perception of Competence and Attitude towards Groups have the highest means. However, it is interesting that the use of two different groups of schools for the Main Study and Control Experiment does not disturb the rank order of mean scores on factors (“Competence”-“Groups”-“Others”-“Outdoor Environment”-“Confidence”), providing some reassurance that events other than a school effect are influencing the outcome.

### **Statistical significance of PSD scores by pupils**

The features of the data reported here have been tested for statistical significance and the results are reported in Table 3.2. The statistical tests were carried out on the pupils’ PSD means, both for the whole sample and for boys and girls and extended to differences between boys and girls. In Appendix 3.3, “Effect sizes and statistical significance of differences in pupils’ mean scores in factors”, these tests are extended further for the five factors in the questionnaire. Non-parametric statistical tests were again used: Wilcoxon Signed-Rank for related samples and the Wilcoxon Rank-Sum and Mann-Whitney Tests for unrelated samples. In the tables below, statistically significant small (0.2) to medium (0.5) effect sizes are shown in bold type.



**Table 3.2** Results and outcomes from statistical tests on PSD mean scores of pupils in both the Main Study and Control Experiment. Upper table, between tests; lower table between genders.

Test	Effect size, $r^*$ Sample size in brackets	2-tailed significance, $p$	Outcome
Main Study, between Pre-test and Post-test	All cases (368): <b>0.21</b> Boys (184): <b>0.20</b> Girls (184): <b>0.22</b>	All cases: 0.000 Boys: 0.000 Girls: 0.000	Post-test pupil means are greater than Pre-test pupil means by a statistically significant difference of small size.
Main Study, between Pre-test and Delayed-test	All cases (308): <b>0.17</b> Boy (160)s: <b>0.22</b> Girls (148): 0.12	All cases: 0.000 Boys: 0.000 Girls: 0.045	Delayed-test pupil means are smaller than Pre-test pupil means by a statistically significant difference of small size for the whole sample and for boys.
Main Study, between Post-test and Delayed-test	All cases (308): <b>0.37</b> Boys (160): <b>0.36</b> Girls (148): <b>0.35</b>	All cases: 0.000 Boys: 0.000 Girls: 0.000	Delayed-test pupil means are smaller than Post-test pupil means by a statistically significant difference of small to medium size.
Control Expt, between Pre-test and Post-test	All cases (117): 0.05 Boys (68): 0.15 Girls (49): 0.12	All cases: 0.465 Boys: 0.083 Girls: 0.233	There is no statistically significant difference between the Pre-test and Post-test pupil means. N.B. boys' scores increase but girls' scores decrease from Pre-test to Post-test.

\* According to Cohen (1988, p.25), a small effect size of 0.2 is difficult to identify amongst extraneous noise.

Test by gender	Effect size, $r$	2-tailed significance, $p$	Outcome
Main Study, Pre-test, between boys (184 cases) and girls (184 cases)	<b>0.16</b>	0.002	Throughout the Main Study girls' means are higher than boys' means by a consistently small but statistically significant amount.
Main Study, Post-test, between boys (184 cases) and girls (184 cases)	<b>0.16</b>	0.002	
Main Study, Delayed-test, between boys (169 cases) and girls (159 cases)	<b>0.16</b>	0.003	
Control Expt, Pre-test, between boys (68 cases) and girls (49 cases)	0.11	0.240	There is no statistically significant difference between boys' and girls' means in the Control Experiment.
Control Expt, Post-test, between boys (68 cases) and girls (49 cases)	0.01	0.886	

Statistical tests were also carried out to establish the significance of the difference in the general level of scoring in the Main Study and the Control Experiment, observed to be 4.0-4.5% lower in the Control Experiment. These were carried out on the Pre-test mean scores in order to avoid any influence from the intervention, which is confined to the Main Study.

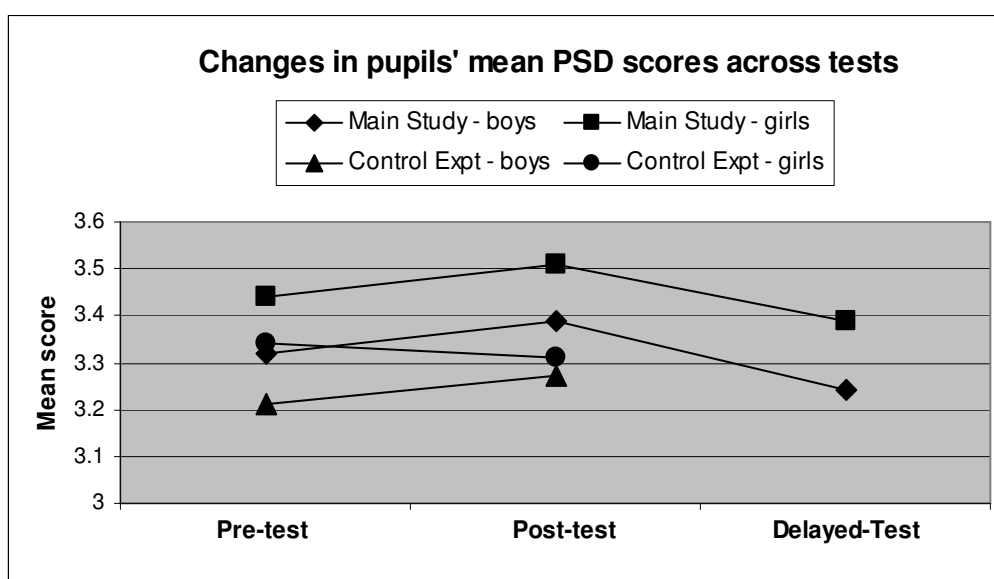
Wilcoxon Rank-Sum and Mann-Whitney Tests for unrelated samples were used. The difference in Pre-test means of 0.14 (c.4.2%), is statistically-significant ( $p=0.001$ ), and yields a small effect size of  $r=0.15$ . For PSD items only, the difference of 0.11 (c.3.3%), is statistically significant ( $p=0.004$ ), but yields only a very small effect size of  $r=0.13$ .

The statistical tests reported here confirm the observations made from the inspection of pupil means. The same tests carried out on the five factors in the questionnaire (Appendix 3.3, “Effect sizes and statistical significance of differences in pupils’ mean scores in factors”) add some detail to the results. We can now make more confident observations on the basis of the statistical outcomes, as follows.

- The Main Study yields coherent scoring from pupils across the three tests, with many statistically-significant effects in PSD scores of small to medium size. On the other hand, within the Control Experiment, which we have noted yields lower scores than the Main Study, there are no statistically significance differences between the Pre-test and the Post-test or between boys and girls. These clear differences in the statistical outcomes of the two experiments raise the possibility of a number of processes at work: a school effect creating a difference between Main Study and Control Experiment scores, which is very small but statistically significant; an anticipation effect around the residential week for the Main Study schools thus elevating Pre- and Post-test scores above Delayed-test and Control Experiment scores; and a real improvement in self-perception of PSD attributes in the Main Study generated by the residential week, albeit lost by the time the Delayed-test was carried out. Further interpretation of these three possibilities will be presented in Chapters 4 and 5.
- In the Main Study the increase in scores from Pre-test to Post-test (+1.8%) and then a strong decrease from Post-test to Delayed-test (-3.3%) is seen to be statistically significant for the whole sample and for both boys and girls. However, broken down by factor, these changes are only significant across statistical tests in Perception of Competence and Awareness of Others, whereas in Attitude towards Groups they are either vanishingly small or not significant. There is little difference in this pattern between boys and girls. In the Control Experiment there are no significant signals at the level of factors. “Competence” also records the highest mean score from pupils and is clearly an area of PSD in which pupils feel most positive about their skills.

- “Competence” and “Others” also stand out as the areas of PSD in which girls out-score boys most emphatically. In the Main Study, at the level of the whole questionnaire, girls score higher than boys in a statistically significant way in all three of Pre-, Post- and Delayed-test; however, broken down by factor, statistical significance is only observed in Perception of Competence and Awareness of Others.
- A final point to make is that while scores for boys and girls generally increase from the Pre-test to the Post-test in both the Main Study and Control Experiment, an exception is that the mean score for girls decreases in the latter. This adverse trend is shown in the Figure 3.4. Statistical tests indicate that the trends in the Control Experiment are not significant, in part because the sample size is low (68 for boys; 49 for girls), but the trends do raise the possibility that for girls the lack of an intervention in that experiment is reflected in the decrease in their score.

**Figure 3.4** Changes in pupils’ mean PSD scores across tests



### **Pupils’ scores on Awareness of the Outdoor Environment items**

The group of six questions on Awareness of the Outdoor Environment was included in the questionnaire in order to discover whether a strong prior awareness made children more

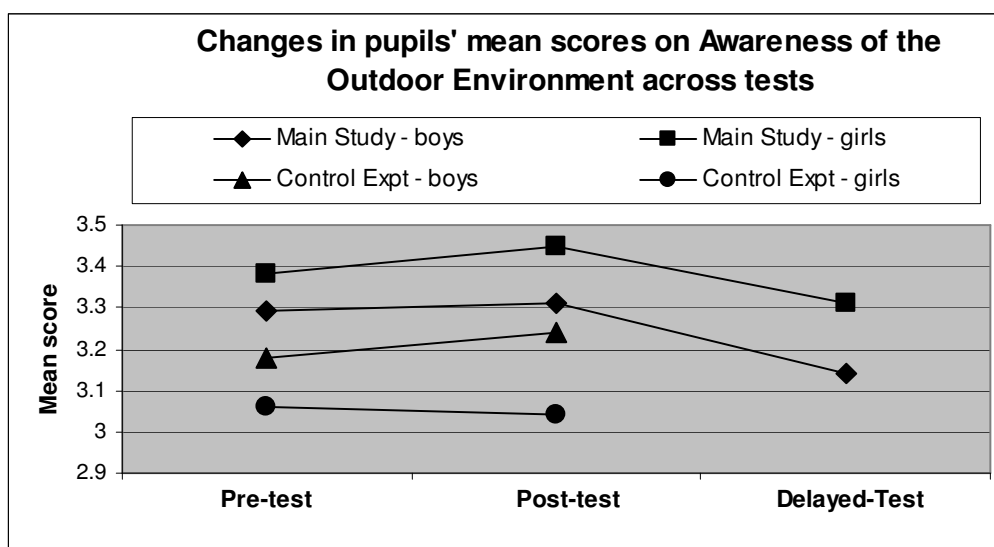
likely to benefit from a residential week of outdoor adventure activities. In the following section on Regression Analysis this question will be addressed directly, but here it would be useful to inspect and test for statistical significance the scores on this factor. The relevant plots and mean scores are given in Appendix 3.2, “Histograms for the distribution of pupils’ means scores in factors”.

Inspection of the histograms for both the Main Study and Control Experiment reveals a tendency for scores to be more widely distributed on this factor than on others. There is a bi-modal distribution in some instances, and peaks in the score count (vertical axis) are lower than in other factors. Mean scores nevertheless fall within the range of those for other factors. Those that might be tested for statistically-significant differences are listed in the following table (Table 3.3). Figure 3.5 is for changes in pupils’ mean scores on Awareness of the Outdoor Environment items across tests and can be compared with Figure 3.4. Girls score higher than boys in the Main Study but, in a clear departure from previous trends, lower than the boys in the Control Experiment. Indeed, mean scores of about 3.05 for girls in the Control Experiment are remarkably low for girls, and approached only by their mean scores in Perception of Confidence in the Control Experiment, raising the possibility that girls at the schools participating in the Control Experiment have a particular lack of awareness of the outdoor environment.

**Table 3.3** Pupils’ mean scores on the factor Awareness of the Outdoor Environment

		<b>All cases</b>	<b>Boys</b>	<b>Girls</b>
Main Study	Pre-test	3.33	3.29	3.38
	Post-test	3.38	3.31	3.45
	Delayed Test	3.22	3.14	3.31
Control Expt	Pre-test	3.13	3.18	3.06
	Post-test	3.16	3.24	3.04

**Figure 3.5** Changes in pupils' mean scores on Awareness of the Outdoor Environment items across tests



The statistical tests conducted on the means from PSD items were also conducted here for “Awareness of the Outdoor Environment” (Table 3.4). Statistically significant small (0.2) to medium (0.5) effect sizes are shown in bold type.

A statistical test was also carried out to establish the significance of the difference in the general level of scoring in the factor in the Main Study and the Control Experiment. The difference in Pre-test means of 0.20 (c.6.2%), is statistically-significant ( $p=0.001$ ), and yields a small effect size of  $r=0.16$ .

The principal observation is that only some of the differences in mean scores are of statistically-significant effect size. Coupled with the wider, even bimodal, distribution of scores, this would suggest that there are more widely held views amongst pupils on their awareness of the outdoor environment. Although this widely held view changed little from before to after the residential week in the Main Study, there was still a significant decline in scores by the time of the Main Study Delayed-test some 2-3 months later (Figure 3.5). On the other hand, the difference in levels of scoring between the two groups of schools used in the Main Study and the Control Experiment is, if anything, stronger. Possible causes of this outcome have already been mentioned above.

**Table 3.4** Results and outcomes from statistical tests on the mean score of pupils on “Awareness of the Outdoor Environment” in both the Main Study and Control Experiment. Upper table, between tests; lower table between genders.

Test	Effect size, $r^*$ Sample size in brackets	2-tailed significance, $p$	Outcome
Main Study, between Pre-test and Post-test	All cases (368): 0.10 Boys (184): 0.05 Girls (184): <b>0.15</b>	All cases: 0.009 Boys: 0.306 Girls: 0.005	For girls, Post-test pupil means are greater than Pre-test pupil means by a statistically significant amount of a small size.
Main Study, between Pre-test and Delayed-test	All cases (308): <b>0.16</b> Boy (160): <b>0.19</b> Girls (148): 0.13	All cases: 0.000 Boys: 0.001 Girls: 0.031	Delayed-test pupil means are smaller than Pre-test item means by a statistically significant difference, of small size for the whole sample and for boys.
Main Study, between Post-test and Delayed-test	All cases (308): <b>0.29</b> Boys (160): <b>0.30</b> Girls (148): <b>0.28</b>	All cases: 0.000 Boys: 0.000 Girls: 0.000	Delayed-test pupil means are smaller than Post-test pupil means by a statistically significant difference of small to medium size.
Control Expt, between Pre-test and Post-test	All cases (117): 0.02 Boys (68): 0.05 Girls (49): 0.03	All cases: 0.758 Boys: 0.524 Girls: 0.793	There is no statistically significant difference between the Pre-test and Post-test pupil means. N.B. boys' scores increase but girls' scores decrease from Pre-test to Post-test.

Test by gender	Effect size, $r$	2-tailed significance, $p$	Outcome
Main Study, Pre-test, between boys (184 cases) and girls (184 cases)	0.07	0.172	In the Main Study girls' means are higher than boys' means by a very small but statistically significant amount in the Post-test and Delayed-test.
Main Study, Post-test, between boys (184 cases) and girls (184 cases)	0.11	0.040	
Main Study, Delayed-test, between boys (169 cases) and girls (159 cases)	0.13	0.015	
Control Expt, Pre-test, between boys (68 cases) and girls (49 cases)	0.15	0.096	There is no statistically significant difference between boys' and girls' means in the Control Experiment.
Control Expt, Post-test, between boys (68 cases) and girls (49 cases)	0.16	0.085	

### **3.4 Regression Analysis**

A regression analysis permits an inspection of the way one score relates to or depends on another score (univariate regression) or several other scores (multivariate regression). For example, “Do pupils who score themselves highly in the Pre-test also score themselves highly in the Post-test?”, thus indicating that the uplift in scores from Pre-test to Post-test, seen in the Main Study, is not due to a random redistribution of scores but due to an orderly uplift across all pupils. Regression is an ideal tool for one of the main objectives of this project, namely to discover how the change in a pupil’s PSD score from before to after a residential week of OAA depends on the pupil’s attributes and interests before the residential week. However, in pursuing this objective it became clear that there are limitations imposed on the relevant data by the scoring process. This will be explained in due course, but in consequence of it, only a limited amount of the regression analysis uses the change in a pupil’s PSD score as the measure of development. Instead, considerable use is made of the pupil’s actual PSD score after the residential week as a measure of “benefit”.

All of the analyses are carried out using pupils’ mean scores as described in the previous section. From time to time comparisons are made between the Main Study and the Control Experiment in order to assess the likely impact of an intervention effect.

#### **3.4.1 Univariate regressions**

Initially, it is informative to inspect the regression between Pre-test, Post-test and Delayed-test scores to establish a baseline against which other regressions can be interpreted. The graphs in Appendix 3.4 show these regressions for the Main Study and Control Experiment for all items, PSD items and items on “Awareness of the Outdoor Environment”, for all cases and for boys and girls.

#### **Univariate regressions between Pre-test, Post-test and Delayed-test scores**

In the Main Study, using all cases, all regressions (all items, PSD items and outdoor environment items) show a good, simple linear model fit to the data, with the independent variable typically sharing about 60% of its variation with the dependent variable ( $R^2 \approx 0.6$ ). An exception to this pattern is on the factor “Awareness of the Outdoor Environment”, where regressions involving the Pre-test data show slightly higher data scatter and an  $R^2$  of

about 0.5. The equivalent factor in the Control Experiment yield  $R^2$  values of only about 0.45. It has been noted before that this factor yields the greatest variability in pupils' scores and, hence, presumably, the greatest divergence in pupils' personal views. Despite differences in the size of the data sets and the degree of scatter, it is clear that pupils who score high on the questionnaire in one test also score high in the other tests, and *vice versa*. It is reassuring to see that pupils remain reasonably consistent in the ranking of their scores on the questions over the period between Pre-test and Post-test, but more reassuring, remain consistent over the 2-3 month period between the Pre- and Post-tests and the Delayed-test in the Main Study, even though pupils' scores fell by a significant amount (about 3%) in the Delayed-test. From this we can deduce that at upper primary school age pupils do have established views that stand the test of time on the environmental and PSD issues in the questionnaire. This is somewhat contrary to one of the arguments put forward for the tailored design of the questionnaire, namely that views are changing rapidly at this stage in life.

Broken down by gender, in the Main Study boys show an almost identical pattern of regressions to that of the whole sample, whereas girls consistently show greater scatter in their mean scores in the regressions of Delayed-test scores on Pre-test scores. In the Control Experiment, boys again show a pattern similar to that of the whole sample, but the girls in this case show stronger  $R^2$  value for Post-test vs Pre-test for all items and PSD items. From these observations we might deduce that girls are more likely than boys to show departures from the general trends in the regressions, either towards weaker correlations (Main Study) or towards stronger correlations (Control Experiment). However, this deduction is tenuous and should simply be noted.

An important feature of all the graphs in Appendix 3.4 is the nature of the gradients of the best-fit straight line. It is observed that the gradient is always less than 1.0 (confirmed in the regression equations), with the high end of the fit passing through or just below the point (4.0, 4.0) and the low end having an intercept on the ordinate of up to +1.2. At the high end, the line is bound to pass close to (4.0, 4.0) because of the upper limit of 4 on scores on the Likert scale and the negatively skewed nature of the datasets. However, at the low end, the positive intercept is indicative of higher scores on the ordinate than on the abscissa, and since the graphs are arranged so that the data plotted on the ordinate postdate the data on the abscissa in terms of their collection, e.g. Post-test follows two weeks after Pre-test, we might



conclude that at least pupils scoring at the low end of the scale are registering higher scores on second and later administrations of the questionnaire.

### **Factors influencing regression results**

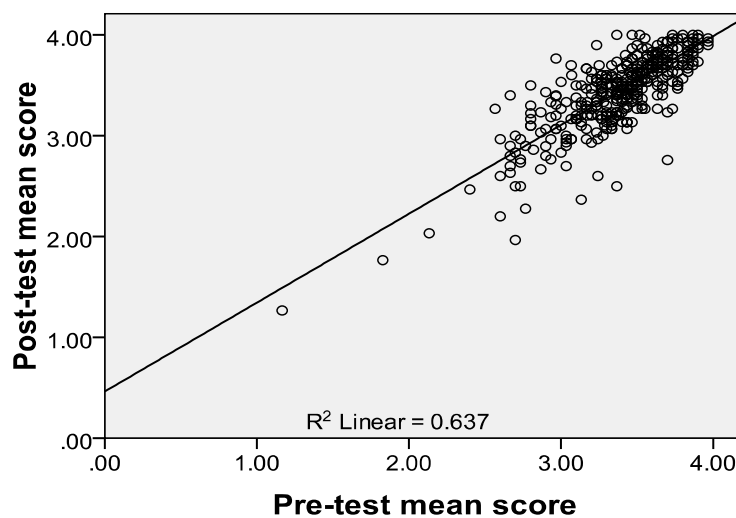
At this point we digress briefly in order to recognise and understand the effect of limitations in the scoring process that impinge on the observation made in the previous section and on the regressions in general. The first limitation relates to a capping effect on the change in pupils' mean scores from before to after a residential week of OAA caused by the 4-point Likert scale used in the questionnaire and the fact that scores are highly skewed towards high values. This effect arises because, in this project, the Likert scale is being used as an interval rather than as a rating scale. The effect is readily seen in Figure 3.6b) in which a scatter plot shows change in pupils' mean scores from Pre-test to Post-test in the Main Study against their Pre-test mean scores. The theoretical limit of this scale capping effect is shown labelled on the plot. For example, if a pupil has a mean of 3.8 in the Pre-test, the maximum benefit s/he can register is 0.2. As a result of this effect, the highest scores in the Post-test might be suppressed to some extent, with a knock-on effect to all observations and analyses that use these scores. Some effect sizes presented in the previous sections of this chapter might be suppressed, such as the effect size for difference between Pre-test means and Post-test means. Having observed this limitation, a pupil-by-pupil analysis of the data shows that only eight of the 368 pupils in the Pre- and Post-tests of the Main Study have a *mean* Post-test score that places their data point on the capping limit. However, a much larger number of pupils will have been capped on their scores on individual questions and, in the absence of a score-by-score analysis of the data, the full extent of capping is unknown.

Is the capping effect a flaw in the use of a Likert scale as an interval scale when data sets are likely to plot towards one end of the scale or the other? Is skewness in such data sets an inevitability? These questions are addressed in Chapter 4, "Implications for research methodology". It is certainly a limitation on the analysis of the data presented here. In our case, assuming the effect is present in all subsets of the whole data set, which is the set illustrated in Figure 3.6, it will affect the quality of regression analyses, including regressions of the change in a pupils' mean score from before to after a residential week of OAA on their Pre-test mean scores. However, there is an additional, second weakness in the dataset that also mitigates against such regressions, and that is the large amount of scatter in

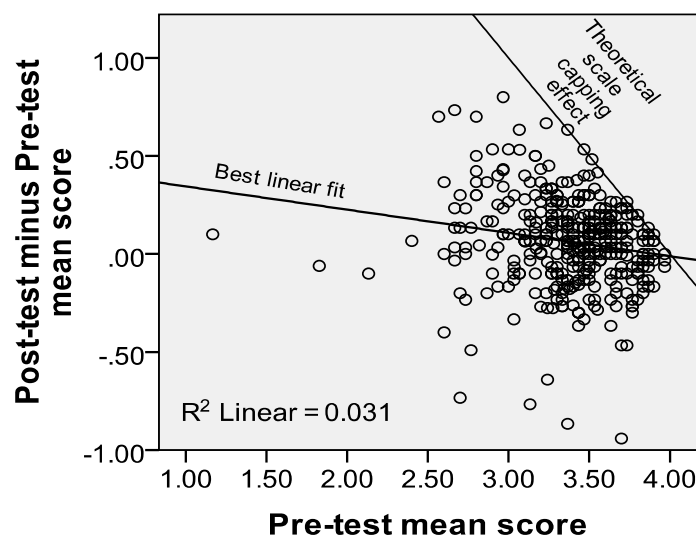
the calculated change in a pupil's PSD score from before to after a residential week of OAA. Figure 3.6b) also shows this: the scatter is so large that it yields an  $R^2$  value of only 0.03.

**Figure 3.6** Univariate linear regressions for the dependence of change in pupils' mean scores between tests in the Main Study, using all cases and all items: a) data used for the regressions in this figure, copied from Appendix 3.4; b) SPSS scatter plot of dependence of change in score Post-test – Pre-test against Pre-test score; c) graph created using one-tenth bands of data to illustrate more clearly the magnitude of the dependence in b), the chained line being the theoretical scale capping effect; d) as for c) but broken down by gender; e) the dependence of the change in score between Post-test and Delayed-test on Post-test score; f) as for e) but broken down by gender.

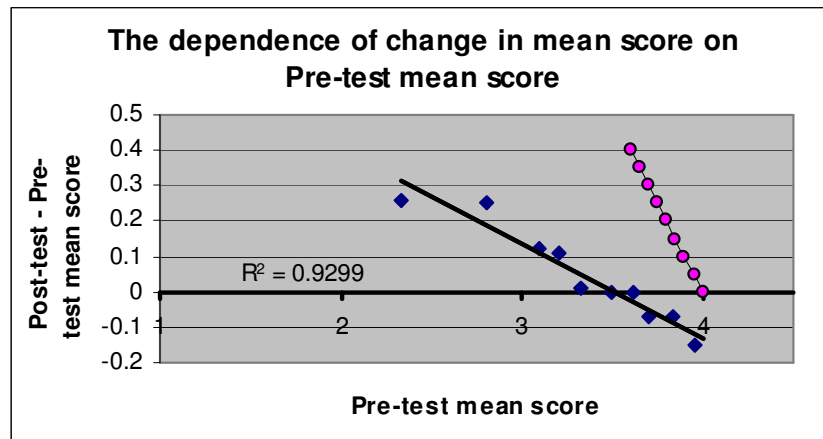
a)



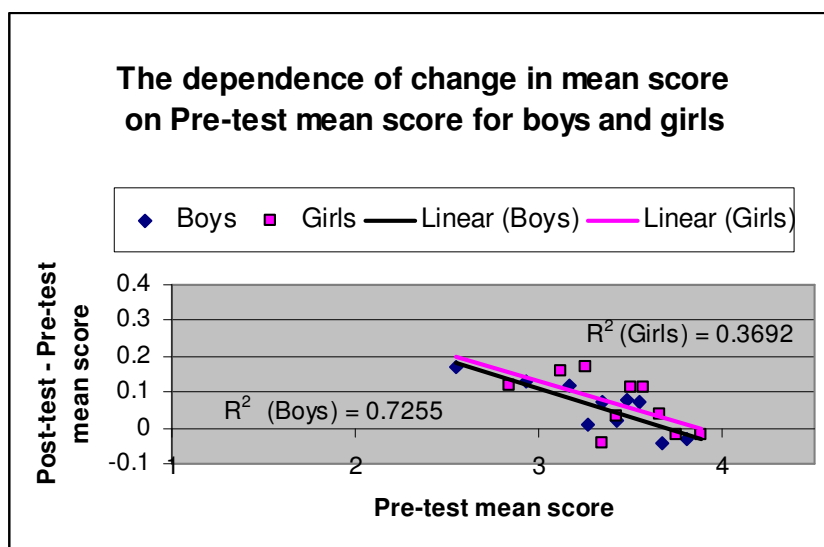
b)



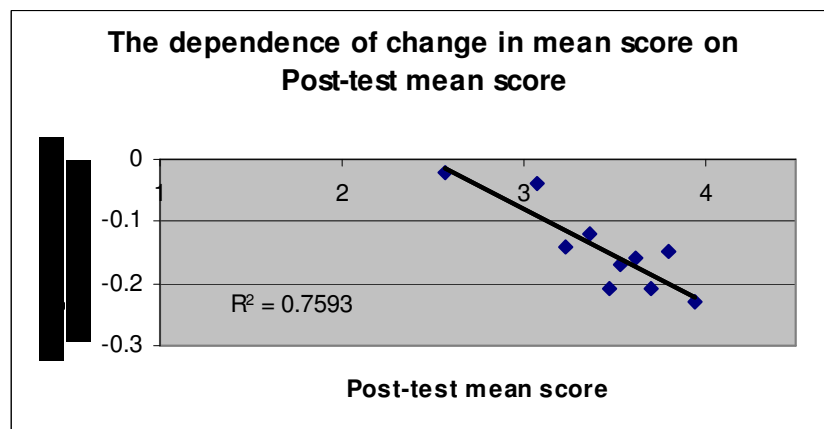
c)



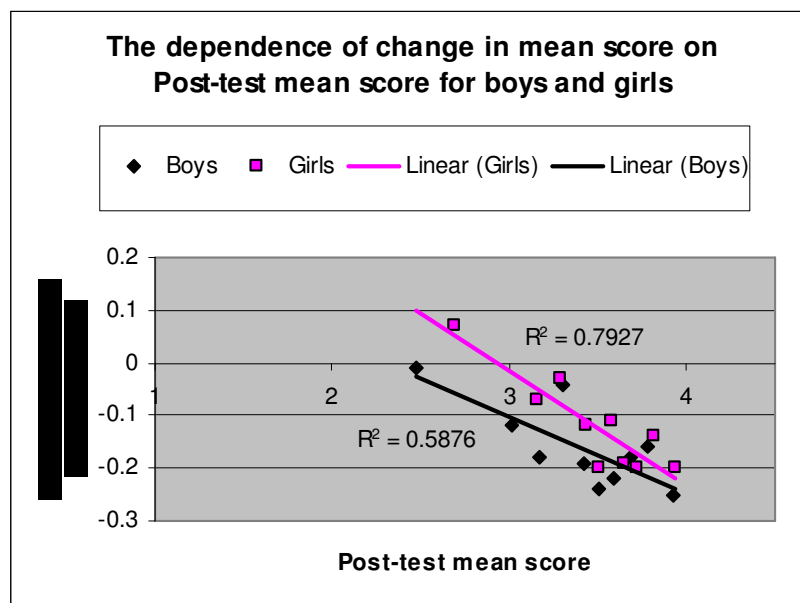
d)



e)



f)



### Avoiding the factors influencing regression results

To at least attempt a regression of the change in pupils' mean scores from before to after a residential week of OAA on their Pre-test mean scores, which is one of the principal objectives of this research, and at the same time avoid the two limitations – scale capping effect and excessive scatter – the data set of all samples was divided into one-tenth subsets from the highest scoring one-tenth in the Pre-test to the lowest scoring one-tenth. The result of the regression using these subsets is shown in Figures 3.6c). A remarkably clear and interesting trend emerges. The trend implies that low-scoring pupils benefit from the residential experience more than high-scoring pupils by as much as 0.3 on the Likert scale, with an  $R^2$  value of 0.93, thus confirming our initial interpretation of the gradients on the graphs in Appendix 3.4, made in the section on univariate regressions between Pre-test, Post-test and Delayed-test scores. Not too much should be read into the negative change in mean scores amongst the high-scoring pupils, however, because that is where the scale capping effect is operating. Figure 3.6d) breaks the data set down by gender. The smaller number of samples per data point is beginning to reflect the large amount of scatter in the whole data set, especially for girls ( $R^2 = 0.37$ ) as opposed to boys ( $R^2 = 0.73$ ), but the underlying trend is still present.

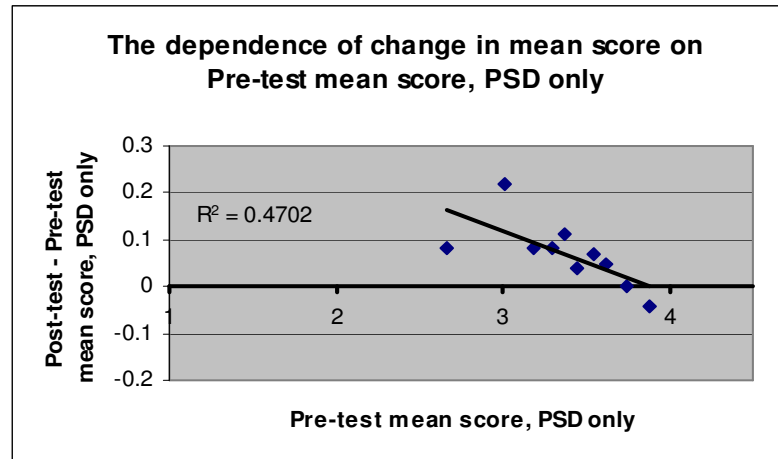
The possibility of low-scoring pupils benefiting more from the residential experience is an attractive one, since OAA has been extensively used to improve self belief amongst less able individuals, e.g. North Lanarkshire “Raising Achievement” programme, for which there was preferential selection of less accomplished pupils (Christie, 2004). However, like so many observations made in this study, there are possible alternative interpretations. In this case, the scale capping effect could be more severe than is thought. Moreover, since the same trend in gradients occurs in the Control Experiment data (Appendix 3.4), where there was no intervention, there is an argument for its cause being an extraneous effect such as scale capping or, alternatively, familiarity with the concepts in the questionnaire influencing the low-scoring pupils more than the high-scoring pupils.

In view of the “success” with the ten subsets of data, this approach to regression was extended to the change in scores between the Post-test and the Delayed-test (Figures 3.6e) & f)). Remember that the Delayed-test registered a decrease in mean scores. The graphs show a trend of low-scoring pupils’ mean scores decreasing by a very small amount whereas those of the high-scoring pupils decreasing by up to 0.2 on the Likert scale ( $R^2 = 0.76$ ). Broken down by gender, the data show that the trend exists for both boys ( $R^2 = 0.59$ ) and girls ( $R^2 = 0.79$ ). In this case, the scale capping effect on the magnitude of change did not operate because the scores were decreasing. As to the cause of the decrease, we might suggest that it is due to teachers not integrating the residential experience into classroom work during subsequent months. However, as mentioned elsewhere, it could be the result of a sense of wellbeing amongst pupils at the time of the residential experience raising their scores in the Pre-test and Post-test above a more “normal” level registered in the Delayed-test.

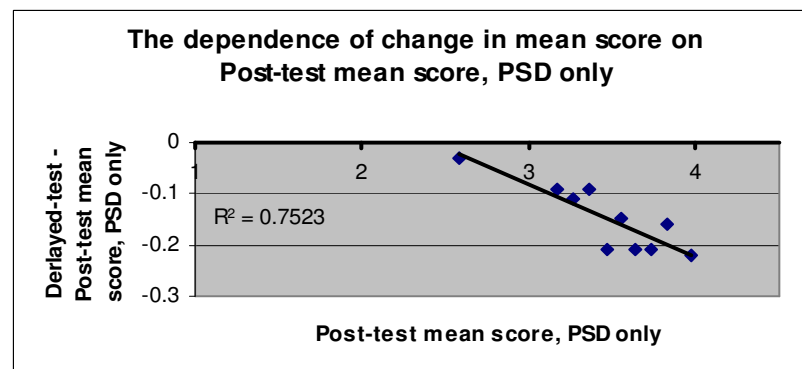
The approach to data analysis of using one-tenth subsets of the data was extended to PSD mean scores only. Regressions of the change in mean scores on PSD items only from Pre-test to Post-test and from Post-test to Delayed-test are presented in Figure 3.7. These are for all pupils only and not broken down by gender. For the dependence of change from before to after the intervention on Pre-test PSD scores and of change from Post-test to Delayed-test on Post-test PSD scores, more-or-less the same trends as we have seen for all items are evident. In addition, a regression of change in PSD score from before to after the intervention on Pre-test “Awareness of the Outdoor Environment” scores has been carried out (Figure 3.7c)) and shows quite clearly that change is not dependent on awareness of the outdoor environment. This is an important conclusion that is confirmed in subsequent regression analyses.

**Figure 3.7** The dependence of change in mean PSD scores between tests in the Main Study: a) the dependence of change in PSD score Post-test – Pre-test on Pre-test PSD score; b) the dependence of the change in PSD score between Post-test and Delayed-test on Post-test PSD score; c) the dependence of change in PSD score Post-test – Pre-test on awareness of the outdoor environment.

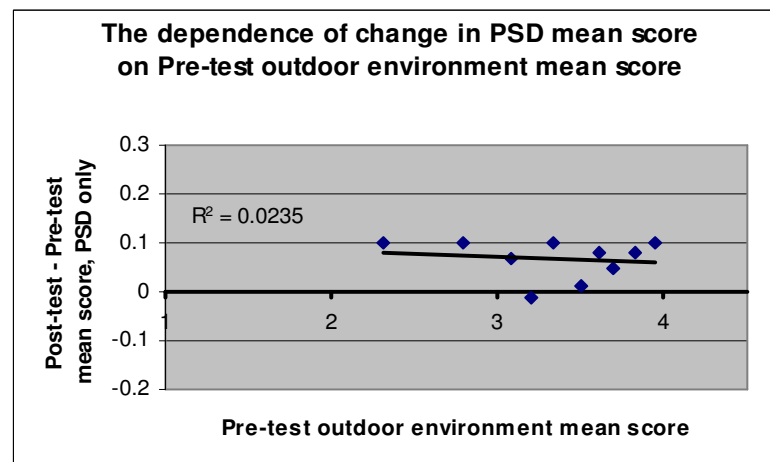
a)



b)



c)



No further attempts were made to carry out regression analysis on the change in pupils' mean scores from before to after their residential week of OAA using all samples in the data set, as illustrated in Figure 3.6b), because of the high level of scatter in the measurements of change. Regressions reported in the rest of this chapter are regressions of Post-test mean scores on a variety of Pre-test mean scores.

### **Univariate regressions using individual PSD factors**

Although not illustrated with graphs in this thesis, univariate regression analyses of the kind illustrated in Appendix 3.4 for the single factor "Awareness of the Outdoor Environment" have also been carried out for all of the separate factors included under the PSD umbrella. Since the importance of individual factors will be considered in the following section on multivariate analysis, the outcome of these analyses in terms of  $R^2$  values is given in the following summary table (Table 3.5). The equivalent  $R^2$  values for "Awareness of the Outdoor Environment" are copied from Appendix 3.4 and included at the bottom of this table for completeness.

Two features of the Main Study are clear from this table. First, there is remarkable uniformity in  $R^2$  values across the various factors, which fall in the range of about 0.4 to 0.5. Second, against this background, some values for girls sit outside this range, reaching down to 0.38 and up to 0.60. This reinforces the observation made above that girls are more likely than boys to show departures from the general trends. Values for the Control Experiment are low, lower than those we see for "Awareness of the Outdoor Environment", and suggest a high degree of scatter in pupils' scores on PSD items in that experiment, in particular boys' scores. This is interpreted as either a school group effect or due to the lack of anticipation of an OAA intervention in the Control group of schools.

Also in anticipation of considering the combined effect of all factors on multivariate regression in the following section, a brief inspection of the potential of any one of the Pre-test factors to predict the outcome in the Post-test score on PSD items is presented here. The table below (Table 3.6) gives  $R^2$  values for the regression of the mean score for Post-test PSD items only on the mean score of each of the five Pre-test factors. Delayed-test data are not included in this analysis.

**Table 3.5** Summary of  $R^2$  values for univariate regressions between tests for pupils' mean scores within PSD factors in the Main Study and Control Experiment.

PSD Factor		$R^2$ values, Main Study			$R^2$ values, Control Expt
		Post-test vs Pre-test	Delayed-test vs Pre-test	Delayed-test vs Post-test	Post-test vs Pre-test
Perception of Competence	All cases	0.48	0.51	0.49	0.39
	Boys	0.45	0.52	0.48	0.39
	Girls	0.50	0.52	0.50	0.37
Perception of Confidence	All cases	0.45	0.41	0.47	0.39
	Boys	0.48	0.42	0.52	0.36
	Girls	0.43	0.41	0.42	0.51
Awareness of Others	All cases	0.48	0.45	0.47	0.33
	Boys	0.43	0.47	0.49	0.28
	Girls	0.50	0.38	0.39	0.44*
Attitude towards Groups	All cases	0.46	0.43	0.50	0.26
	Boys	0.50	0.43	0.44	0.23
	Girls	0.38	0.41	0.60	0.37

Awareness of the Outdoor Environment	All cases	0.53	0.48	0.61	0.43
	Boys	0.50	0.53	0.64	0.41
	Girls	0.55	0.38*	0.54	0.45

\* Not significant, otherwise all regressions statistically significant.

Two observations can be made on these values. First, data in the Control Experiment yield poor  $R^2$  values, reflecting a high level of scatter in the data which, to say again, might relate to school group or to the absence of an intervention associated with this experiment. Second, in the Main Study, “Awareness of the Outdoor Environment” taken on its own is a particularly poor predictor of Post-test PSD scores. Other factors perform better, but none is very good at predicting PSD scores after the intervention. The following section on multivariate regression considers the potential of combining Pre-test factors to predict Post-test PSD scores in the Main Study.

### Summary of Section 3.4.1

It should be pointed out that other than making use of the  $R^2$  values generated in SPSS and checking that regression parameters are statistically significant, no statistical analysis of the *quality* of the univariate regressions presented above, e.g. distribution of residuals, has been carried out. Nevertheless, the main findings from section 3.4.1 are clear:



**Table 3.6** Summary of  $R^2$  values for univariate regression of pupils' mean Post-test PSD scores on their mean scores in individual factors in the Pre-test, for both the Main Study and Control Experiment.

Pretest predictor	$R^2$ values for Post-test PSD items only					
	Main Study			Control Experiment		
	Whole sample	Boys	Girls	Whole sample	Boys	Girls
Awareness of the Outdoor Environment	0.21	0.27	0.12	0.22	0.24	0.22
Perception of Competence	0.47	0.47	0.45	0.28	0.28	0.29
Perception of Confidence	0.43	0.43	0.44	0.34	0.29	0.46
Awareness of Others	0.44	0.41	0.44	0.31	0.28	0.39
Attitude towards Groups	0.41	0.48	0.32	0.27	0.23	0.35

All regressions statistically significant.

- pupils who score high on the questionnaire in one test also score high in the other tests, and *vice versa*, albeit with considerable scatter in some cases;
- the ceiling on scores at the top end of the Likert scale might distort the mean scores on tests, but, equally, there appears to be a pupil effect that indicates those that score low benefit the most from their residential experience and tend to retain that benefit;
- Awareness of the Outdoor Environment appears to be a particularly poor predictor of any PSD benefits experienced by pupils, whereas individual PSD factors are all somewhat better in their ability to predict the Post-test PSD mean score of pupils.

### 3.4.2 Multivariate regressions

This section addresses a core objective of this project, which is to discover how pupils' prior personal and social skills and their awareness of the outdoor environment might predict pupils' PSD during a residential week of outdoor adventure activities. Ideally, the measure of benefit used here would be the change in scores on PSD items in the Main Study from Pre-test to Post-test and its retention in the Delayed-test. However, we can see in Figure 3.6b) above that in regressions involving the small changes of score recorded, data scatter is high and  $R^2$  values are vanishingly small (0.03 in Figure 3.6b)). Indeed, attempts to model change of score against Pre-test score yield poor regression results. It was decided, therefore, to make Post-test PSD mean score the target variable as a measure of the benefit of the residential week. This provides much better quality regression results, and although it is not a measure of the change over the week, we now know from using one-tenth subsets of the data and the discussion in the previous section something about the changes that might

take place, i.e. that high-scoring pupils show little or no change whereas low-scoring pupils show positive change and retain it.

## Initial regression analysis

This multivariate regression analysis takes the so-called “enter” approach of SPSS.

There is little, if any, published research to guide the choice of pupil characteristics that are likely to be the best predictors of Post-test PSD mean score, although the five factors in the questionnaire in use here were chosen to represent the spectrum of characteristics most commonly addressed in the research literature. It was decided, therefore, to begin regression modelling at the most general level. Working with data from the Main Study (368 cases), a regression based on all cases was carried out using the SPSS default option for multivariate linear regression which does not extend priority to any predictor (independent) variable. In the following equation, and all regression equations in this section of the thesis, the symbols used are:

P-t PSD	Post-test mean score on the 24 PSD items
OE	Pre-test mean score on the 6 “Awareness of the Outdoor Environment” items
Comp	Pre-test mean score on the 6 “Perception of Competence” items
Conf	Pre-test mean score on the 6 “Perception of Confidence” items
Others	Pre-test mean score on the 6 “Awareness of Others” items
Groups	Pre-test mean score on the 6 “Attitude towards Groups” items

The initial model is,

$$\text{P-t PSD} = 0.388 + 0.051 \text{ OE} + 0.227 \text{ Comp} + 0.201 \text{ Conf} + 0.185 \text{ Others} + 0.239 \text{ Groups}$$

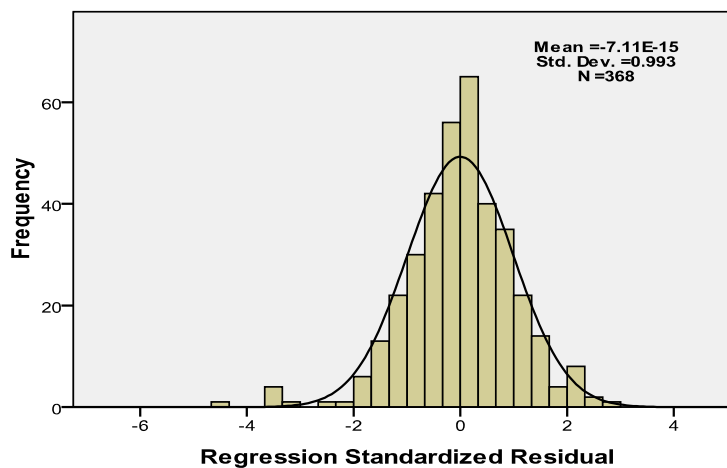
The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.641, i.e. the model accounts for 64.1% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.636, an adjustment of only 0.5%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised coefficients		Standardised coefficients	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	<b>β</b>		
Constant	0.388	0.127		3.047	0.002
Outdoor/Env	0.051	0.031	0.063	1.647	0.100
Comp	0.227	0.049	0.223	4.610	<0.001
Conf	0.201	0.037	0.236	5.488	<0.001
Others	0.185	0.037	0.221	4.979	<0.001
Groups	0.239	0.041	0.247	5.865	<0.001

- Standardised residuals are distributed normally: mean =  $-7.11 \times 10^{-15}$ ; SD = 0.993 (Figure 3.8).

**Figure 3.8** Distribution of standardised residuals from the multivariate regression for Post-test mean PSD score. Note that in other regressions reported here this distribution-of-residuals plot is not illustrated unless it demonstrates an abnormal feature of the quality of the regression.



These results confirm those obtained from the univariate linear regressions in the previous section. “Awareness of the Outdoor Environment” is a poor predictor of PSD outcome whereas the four PSD factors contribute more-or-less equally as predictors. In terms of the actual contribution from each Pre-test factor, we can examine in a little more detail the contribution of “Perception of Competence” as an example of the magnitudes involved. A higher Pre-test mean score on the six competence questions of 0.1 would lead to a higher Post-test PSD mean score of  $0.1 \times 0.227 \approx 0.023$ , other variables being kept constant. We have seen that in the Main Study dataset a statistically significant difference in mean PSD score might be 0.1 (say, 3.3 as opposed to 3.2 on the 4 point Likert scale). This would require a difference of about 0.4 in the Pre-test competence score, which is just within the range of scores from our cases.

The clear lack of importance of “Awareness in the Outdoor Environment” as a predictor of outcome is interesting. Although pupils at late primary stage have classes in environmental studies, the teachers associated with this project made it clear that in the classroom few, if any, connections are made between those studies and the residential week. The residential week is treated very much as a fun week for affective PSD with little pressure on pupils to reap a cognitive reward from it. On the other hand, the children clearly like being outdoors, delivering some of the highest item scores on Q1 and Q11, which ask about “the outdoors”. We might surmise that at age 10-12 years old a connection is not necessarily made between “the outdoors” and the feeling of well-being it produces, and, in particular, any connection between the natural environment, as a facet of “the outdoors”, with such a feeling.

Multivariate regressions without priority for any predictor were carried out for boys and girls separately using Main Study data (a sample size of 184 for each). The modeling results were:

- Boys

$$P\text{-}t \text{ PSD} = 0.386 + 0.075 \text{ OE} + 0.180 \text{ Comp} + 0.181 \text{ Conf} + 0.160 \text{ Others} + 0.304 \text{ Groups}$$

The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.644, i.e. the model accounts for 64.4% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.634, an adjustment of 1.0%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.386	0.171		2.253	0.025
Outdoor/Env	0.075	0.044	0.096	1.706	0.090
Comp	0.180	0.071	0.180	2.543	0.012
Conf	0.181	0.056	0.204	3.239	0.001
Others	0.160	0.054	0.188	2.991	0.003
Groups	0.304	0.061	0.313	4.991	<0.001

- Standardised residuals are distributed normally: mean =  $-1.55 \times 10^{-15}$ ; SD = 0.986

- Girls

$$P\text{-}t \text{ PSD} = 0.451 + 0.010 \text{ OE} + 0.280 \text{ Comp} + 0.222 \text{ Conf} + 0.208 \text{ Others} + 0.167 \text{ Groups}$$

The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.628, i.e. the model accounts for 62.8% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.617, an adjustment of 1.1%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.451	0.202		2.235	0.027
Outdoor/Env	0.010	0.045	0.012	0.277	0.820
Comp	0.280	0.070	0.268	4.012	<0.001
Conf	0.222	0.050	0.278	4.477	<0.001
Others	0.208	0.055	0.243	3.778	<0.001
Groups	0.167	0.056	0.176	2.967	0.003

- Standardised residuals are distributed normally: mean =  $-6.45 \times 10^{-15}$ ; SD = 0.986

Some gender differences emerge from this analysis. For boys, “Attitude towards Groups” emerges as the strongest predictor of Post-test PSD score, whereas for girls this factor is less important than the other three Pre-test PSD factors. With “Awareness of the Outdoor Environment” continuing to be unimportant, the outcomes for boys and girls complement one another to deliver the result seen in the regression for all cases.

## Next-step regression analysis

SPSS provides another approach to entering data into a regression without recourse to pre-judgements on the part of the researcher of the relative importance of the predicting variables. This is the so-called “stepwise” entry of predicting variables, in which the predictor variables are added to the regression equation in the order of their importance according to purely quantitative acceptance criteria established by trial-and-error testing within the software. We might expect the “Awareness of the Outdoor Environment” predictor to be rejected.

Using stepwise entry of all predictors, the following equation is obtained using all cases in the Main Study. As predicted, “Awareness of the Outdoor Environment” has been rejected from this.

$$P\text{-}t \text{ PSD} = 0.442 + 0.249 \text{ Comp} + 0.200 \text{ Others} + 0.209 \text{ Conf} + 0.229 \text{ Groups}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	47.3%
Others	9.1%
Conf	4.2%
Groups	3.2%
Total	63.9%

The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.639, i.e. the model accounts for 63.9% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.635, an adjustment of 0.4%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
<b>Constant</b>	0.442	0.123		3.576	<0.001
<b>Outdoor/Env</b>	excluded				
<b>Comp</b>	0.249	0.047	0.246	5.260	<0.001
<b>Conf</b>	0.209	0.036	0.245	5.736	<0.001
<b>Others</b>	0.200	0.036	0.239	5.524	<0.001
<b>Groups</b>	0.229	0.040	0.237	5.667	<0.001

- Standardised residuals are distributed normally: mean =  $-7.16 \times 10^{-15}$ ; SD = 0.995

In terms of the regression coefficients, this result is very much the same as the one derived from straightforward “entry” of the predictor variables. However, the acceptance criteria for the predictor variables have found Competence to be the strongest predictor. Not too much should be read into the actual values of contributions to  $R^2$ , since even a small amount of collinearity between variables will allow Competence to “steal” some percentage from other variables. It is expected that there will be some collinearity between PSD factors, but tests provided within SPSS for its significance show that it is not damaging to the regression.

Analysing by gender, the results were as follows:

- Boys:

$$P\text{-}t \text{ PSD} = 0.435 + 0.300 \text{ Groups} + 0.193 \text{ Conf} + 0.187 \text{ Others} + 0.205 \text{ Comp}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Groups	47.8%
Conf	10.1%
Others	4.2%
Comp	1.8%
Total	63.8%

The model meets a wide range of criteria for medium to good quality, some of which are as follows:

- $R^2$  is 0.638, i.e. the model accounts for 63.8% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.630, an adjustment of 0.8%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.435	0.170		2.557	0.011
Outdoor/Env	excluded				
Comp	0.205	0.069	0.206	2.956	0.004
Conf	0.193	0.056	0.218	3.460	0.001
Others	0.187	0.052	0.220	3.624	<0.001
Groups	0.300	0.061	0.309	4.903	<0.001

- Standardised residuals are distributed normally: mean =  $-2.64 \times 10^{-15}$ ; SD = 0.989

- Girls:

$$P\text{-}t \text{ PSD} = 0.468 + 0.286 \text{ Comp} + 0.210 \text{ Others} + 0.223 \text{ Conf} + 0.164 \text{ Groups}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	45.2%
Others	11.0%
Conf	4.7%
Groups	1.9%
Total	62.8%

The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.628, i.e. the model accounts for 62.8% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.620, an adjustment of 0.8%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.468	0.188		2.487	0.014
Outdoor/Env	excluded				
Comp	0.286	0.066	0.273	4.345	<0.001
Conf	0.223	0.049	0.279	4.559	<0.001
Others	0.210	0.054	0.246	3.894	<0.001
Groups	0.164	0.054	0.172	3.036	0.003

- Standardised residuals are distributed normally: mean =  $-4.63 \times 10^{-15}$ ; SD = 0.989

The difference between boys and girls is interesting, in that the order of importance of predictors is quite the opposite in the two groups. This echoes the results from the univariate regression analysis and adds some detail to the results from the previous, “entry” approach. In summary, boys attitude towards group work is the best predictor of their perception of their PSD skills following the week of OAA, whereas for girls it is their perception of their competence that best predicts their outcome. The further implication of these results is that scores in the less important variables changed more from before to after the intervention, yielding smaller  $R^2$  values.

A comparative analysis of regressions was carried out for the Control Experiment data as well as the Main Study data but with poor quality results, although both “entry” and



“stepwise” approaches indicate that the order of importance of Pre-test variables is “Perception of Confidence” – “Attitude towards Groups” – “Awareness of the Outdoor Environment”. There are very few similarities in the results of the two experiments.

### **Hierarchical regression analysis**

The “enter” and “stepwise” approaches to entering the predictor variables into the regression rely entirely on mathematical methods to determine the relative importance of the variables. The “hierarchical” approach allows the researcher to determine the order in which variables are entered into the regression equation, entering them in what he/she feels is their order of importance. If more than one variable is entered at any stage, their relative importance is established using the “enter” approach. On the basis of the analyses so far, and with little other evidence to use, the following hierarchies have been adopted. “Awareness of the Outdoor Environment” has not been entered as a variable at all.

- For all cases in the Main Study the four remaining predictor variables have been entered at the same time, i.e. they are given equal importance.
- For boys, “Attitude towards Groups” has been entered first, and then the other three predictors entered second, all at the same time.
- For girls, “Perception of Competence” has been entered first, followed by “Perception of Confidence” and “Awareness of Others” equal second, and finally “Attitude towards Groups”.

Since these hierarchies include all the variables retained in the stepwise approach it is not surprising that the final regression equations are the same as the equations arrived at in the previous section. However, because the “enter” method is used at each hierarchical stage, percentage contributions to  $R^2$  are summed at each stage. The equations are repeated here but without the accompanying notes on quality.

- All cases:

$$P\text{-}t \text{ PSD} = 0.442 + 0.249 \text{ Comp} + 0.200 \text{ Others} + 0.209 \text{ Conf} + 0.229 \text{ Groups}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	47.3%
Others	9.1%
Conf	4.2%
Groups	3.2%
Total	63.9%

- Boys:

$$P\text{-}t \text{ PSD} = 0.435 + 0.205 \text{ Comp} + 0.193 \text{ Conf} + 0.187 \text{ Others} + 0.300 \text{ Groups}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Groups	47.8%
Comp + Conf + Others	16.0%
Total	63.8%

- Girls

$$P\text{-}t \text{ PSD} = 0.468 + 0.286 \text{ Comp} + 0.210 \text{ Others} + 0.223 \text{ Conf} + 0.164 \text{ Groups}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	45.2%
Others + Conf	15.7%
Groups	1.9%
Total	62.8%

### Summary of Section 3.4.2

The regression analyses, both univariate and multivariate, converge on the result that Pre-test score on Attitude towards Groups is the best predictors of what a boy will score in the post-intervention test, whereas Pre-test scores on Perception of Competence followed by Awareness of Others and Perception of Confidence are the best predictors of what a girl will score in the post-intervention test. An Awareness of the Outdoor Environment is not a predictor for either boys or girls.

### 3.4.3 Extension of regression analysis to the Main Study Delayed-test data

This section dispenses with univariate analysis and contains only multivariate analyses of the predictive relationship between Post-test mean scores and Delayed-test PSD mean scores.

The delay of about 2-3 months from the time of the Post-test to the time of the Delayed-test saw a statistically-significant fall in pupils' mean PSD scores of small to medium effect size (see section 3.3.2). Analysed by gender, this fall is seen to occur for both boys and girls, although boys mean scores are consistently lower than girls (Figure 3.6). In section 3.4.1 on univariate linear regressions, it became clear that when the data are analysed by the level at which pupils score, the lower-scoring tenths actually lose very little of their Post-test gains in the Delayed-test while the high-scoring tenths are responsible for the fall in overall mean score (Figure 3.6e)), and this is the case for both boys and girls (Figure 3.6f)).

With regression analysis it is possible to investigate the relationship between scores on the five questionnaire factors at the Post-test stage and the overall PSD score in the Delayed-test, in much the same ways as we investigated the Pre-test – Post-test relationship. In this case, it would be of interest to know what post-intervention characteristics of pupils are key to retaining Post-test gains. In view of the summary in the preceding paragraph, it was thought that this investigation should proceed using the dataset subdivided into equal numbers of high-scoring and low-scoring pupils (154 each) rather than boys and girls. However, this approach only worked in part, because when the high-score data alone are examined, they form an equidimensional cloud of data points yielding very poor regression results. The tail to low scores forms a much better dataset for regression purposes and is reported in what follows. In view of this limitation, the analysis for boys and girls separately was carried out for the whole dataset, since sample sizes would have been too small (c.70-80 cases) for an acceptable regression using only the low-scoring dataset, given the scatter in the data.

### Initial regression analysis

Symbols used here are:

D-t PSD	Delayed-test mean score on the 24 PSD items
OE	Post-test mean score on the 6 “Awareness of the Outdoor Environment” items
Comp	Post-test mean score on the 6 “Perception of Competence” items
Conf	Post-test mean score on the 6 “Perception of Confidence” items
Others	Post-test mean score on the 6 “Awareness of Others” items
Groups	Post-test mean score on the 6 “Attitude towards Groups” items

The initial model using the SPSS “enter” facility for all cases is:

$$\text{D-t PSD} = 0.287 + 0.050 \text{ OE} + 0.284 \text{ Comp} + 0.131 \text{ Conf} + 0.223 \text{ Others} + 0.183 \text{ Groups}$$

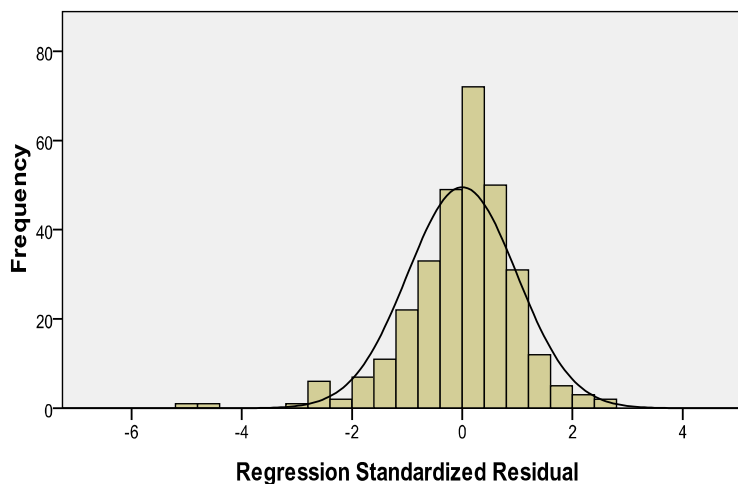
The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.622, i.e. the model accounts for 62.2% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.616, an adjustment of only 1.0%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised coefficients		Standardised coefficients	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.287	0.143		2.006	0.046
Outdoor/Env	0.050	0.037	0.061	1.355	0.176
Comp	0.284	0.059	0.278	4.848	<0.001
Conf	0.131	0.050	0.152	2.607	0.010
Others	0.223	0.056	0.244	4.004	<0.001
Groups	0.183	0.050	0.187	3.665	<0.001

- Standardised residuals are distributed normally: mean =  $1.52 \times 10^{-15}$ ; SD = 0.992. (Figure 3.7).

**Figure 3.9** Distribution of standardised residuals from the regression for Delayed-test mean PSD score. Note that in other regressions reported in this section the distribution-of-residuals plot is not illustrated unless it demonstrates a particular feature of the quality of the regression.



The initial model for lowest-scoring 50% of all cases is:

$$D\text{-}t \text{ PSD} = 0.295 + 0.035 \text{ OE} + 0.355 \text{ Comp} + 0.067 \text{ Conf} + 0.176 \text{ Others} + 0.225 \text{ Groups}$$

The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.570, i.e. the model accounts for 57.0% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.555, an adjustment of 2.6%, therefore the model from the sample generalises up to the whole population only moderately well.
- Significance of coefficients:

	Unstandardised coefficients		Standardised coefficients	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.295	0.204		1.446	0.150
Outdoor/Env	0.035	0.047	0.047	0.750	0.455
Comp	0.355	0.071	0.378	4.967	<0.001
Conf	0.067	0.069	0.072	0.971	0.333
Others	0.176	0.071	0.191	2.483	<0.014
Groups	0.225	0.063	0.248	3.604	<0.001

- Standardised residuals are distributed normally: mean =  $1.75 \times 10^{-15}$ ; SD = 0.984.

The initial model for all boys is:

$$D\text{-}t \text{ PSD} = 0.272 + 0.112 \text{ OE} + 0.274 \text{ Comp} + 0.223 \text{ Conf} + 0.166 \text{ Others} + 0.099 \text{ Groups}$$

The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.603, i.e. the model accounts for 60.3% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.591, an adjustment of 2.0%, therefore the model from the sample generalises up to the whole population reasonably well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.272	0.205		1.325	0.187
Outdoor/Env	0.112	0.053	0.140	2.118	0.036
Comp	0.274	0.093	0.252	2.950	0.004
Conf	0.223	0.078	0.241	2.853	0.005
Others	0.166	0.096	0.172	1.721	0.087
Groups	0.099	0.075	0.104	1.325	0.187

- Standardised residuals are distributed normally: mean =  $4.1 \times 10^{-15}$ ; SD = 0.984

The initial model for all girls is:

$$D\text{-}t \text{ PSD} = 0.311 - 0.035 \text{ OE} + 0.305 \text{ Comp} + 0.063 \text{ Conf} + 0.238 \text{ Others} + 0.294 \text{ Groups}$$

The model meets a wide range of criteria for good quality, some of which are as follows:

- $R^2$  is 0.667, i.e. the model accounts for 66.7% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.656, an adjustment of 1.6%, therefore the model from the sample generalises up to the whole population moderately well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.311	0.202		1.539	0.126
Outdoor/Env	-0.035	0.050	-0.042	-0.699	0.486
Comp	0.305	0.070	0.331	4.354	<0.001
Conf	0.063	0.063	0.082	0.996	0.321
Others	0.238	0.063	0.279	3.778	<0.001
Groups	0.294	0.064	0.298	4.580	<0.001

- Standardised residuals are distributed normally: mean =  $-2.94 \times 10^{-15}$ ; SD = 0.983

### Next step regression analysis

Using stepwise entry of all predictors, the following equation is obtained using all cases in the Main Study. “Awareness of the Outdoor Environment” has been rejected from this equation, but it will be seen in the other equations in this section that it is not always this factor or this factor alone that has been rejected by the criteria established in the software.

$$D\text{-}t \text{ PSD} = 0.319 + 0.245 \text{ Others} + 0.303 \text{ Comp} + 0.183 \text{ Groups} + 0.129 \text{ Conf}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Others	49.9%
Comp	9.2%
Groups	2.1%
Conf	0.8%
Total	62.0%

The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.620, i.e. the model accounts for 62.0% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.615, an adjustment of 0.8%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	$t$	Significance of $t$ -test
	<b>b</b>	Std error	$\beta$		
Constant	0.319	0.141		2.255	0.025
Outdoor/Env	excluded				
Comp	0.303	0.057	0.296	5.312	<0.001
Conf	0.129	0.050	0.150	2.567	0.011
Others	0.245	0.050	0.268	4.610	<0.001
Groups	0.183	0.053	0.187	3.654	<0.001

- Standardised residuals are distributed normally: mean =  $1.63 \times 10^{-15}$ ; SD = 0.993

The stepwise entry model for lowest-scoring 50% of cases is:

$$D\text{-}t \text{ PSD} = 0.356 + 0.392 \text{ Comp} + 0.231 \text{ Groups} + 0.209 \text{ Others}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	44.9%
Groups	8.7%
Others	3.0%
Total	56.6%

The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.566, i.e. the model accounts for 56.6% of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.557, an adjustment of 1.6%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	$t$	Significance of $t$ -test
	<b>b</b>	Std error	$\beta$		
Constant	0.336	0.189		1.825	0.070
Outdoor/Env	excluded				
Comp	0.392	0.063	0.418	6.244	<0.001
Conf	excluded				
Others	0.209	0.064	0.227	3.244	<0.001
Groups	0.231	0.062	0.254	3.731	0.001

- Standardised residuals are distributed normally: mean =  $-1.14 \times 10^{-15}$ ; SD = 0.990

- All boys

$$D\text{-}t \text{ PSD} = 0.335 + 0.217 \text{ Others} + 0.293 \text{ Comp} + 0.235 \text{ Conf} + 0.112 \text{ OE}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Others	49.5%	
Comp	7.1%	
Conf	2.2%	
OE	1.2%	
Total	59.9%	(NB rounding error)

The model meets a wide range of criteria for medium quality, some of which are as follows:

- $R^2$  is 0.599, i.e. the model accounts for 59.9 % of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.589, an adjustment of 1.7%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.335	0.200		1.676	0.096
Outdoor/Env	0.112	0.053	0.140	2.115	0.036
Comp	0.293	0.092	0.270	3.199	0.002
Conf	0.235	0.078	0.254	3.024	0.003
Others	0.217	0.088	0.226	2.459	0.015
Groups	excluded				

- Standardised residuals are distributed normally: mean =  $4.01 \times 10^{-15}$ ; SD = 0.987

- All girls

$$D\text{-}t \text{ PSD} = 0.233 + 0.318 \text{ Comp} + 0.311 \text{ Groups} + 0.254 \text{ Others}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	49.5%
Groups	12.0%
Others	4.9%
Total	66.4%

The model meets a wide range of criteria for good quality, some of which are as follows:



- $R^2$  is 0.664, i.e. the model accounts for 66.4 % of the variance in the Post-test mean score. Adjusted  $R^2$  is 0.657, an adjustment of only 1.0%, therefore the model from the sample generalises up to the whole population very well.
- Significance of coefficients:

	Unstandardised		Standardised	<i>t</i>	Significance of <i>t</i> -test
	<b>b</b>	Std error	$\beta$		
Constant	0.223	0.190		1.226	0.222
Outdoor/Env	excluded				
Comp	0.318	0.059	0.345	5.374	<0.001
Conf	excluded				
Others	0.254	0.055	0.298	4.589	<0.001
Groups	0.311	0.062	0.314	4.983	<0.001

- Standardised residuals are distributed normally: mean =  $-1.74 \times 10^{-15}$ ; SD = 0.990

### Hierarchical regression analysis

Following the “entry” and “stepwise” approaches to entering the predictor variables into the regression, for the “hierarchical” approach the following hierarchies have been adopted.

- For all cases, of the four remaining predictor variables after taking the stepwise approach, “Awareness of Others” and “Perception of Competence” have been entered first, and “Attitude towards Groups” and “Perception of Confidence” have been entered second.
- For the lowest-scoring 50% of cases, of the three remaining variables it is clear that “Perception of Competence” is the strongest predictor, and it has been entered first, with “Awareness of Others” and “Attitude towards Groups” entered second.
- For all boys, there is little to choose in importance between “Perception of Competence”, “Perception of Confidence” and “Awareness of Others”, and so all of these were entered first, with “Awareness of the Outdoor Environment” following.
- For all girls, “Perception of Competence” and “Attitude towards Groups” are marginally stronger predictors and were entered before “Awareness of Others”.

The results were as follows.

- All cases. Here, the regression equation is the same as that from the stepwise approach and the quality measures are the same, but the percentages of variance explained by the predictors have been summed for each of the chosen hierarchies.

$$D-t \text{ PSD} = 0.319 + 0.245 \text{ Others} + 0.303 \text{ Comp} + 0.183 \text{ Groups} + 0.129 \text{ Conf}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Others + Comp	59.1%
Groups + Conf	2.9%
Total	62.0%

- For the lowest-scoring 50% of cases. The same comments apply here.

$$D-t \text{ PSD} = 0.356 + 0.392 \text{ Comp} + 0.231 \text{ Groups} + 0.209 \text{ Others}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Comp	44.9%
Groups + Others	11.7%
Total	56.6%

- All boys

$$D-t \text{ PSD} = 0.335 + 0.217 \text{ Others} + 0.293 \text{ Comp} + 0.235 \text{ Conf} + 0.112 \text{ OE}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Others + Comp + Conf	48.7%
OE	1.2%
Total	59.9%

- All girls

$$D-t \text{ PSD} = 0.233 + 0.318 \text{ Comp} + 0.311 \text{ Groups} + 0.254 \text{ Others}$$

Contributions to the coefficient of determination,  $R^2$ , are as follows:

Groups + Comp	61.5%
Others	4.9%
Total	66.4%

### Summary of Section 3.4.3

The selection of hierarchies in the last section provides a summary of the essential results from extending the regression analysis of the Main Study data to the Delayed-test. It should be noted that although scores decreased from the Post-test to the Delayed-test, as opposed to an increase from Pre-test to Post-test, the correlations revealed here are interpreted in just the same way – for a good predictor, a relatively high score on that predictor in the Post-test tends to lead to a relatively high PSD score in the Delayed test, and *vice versa*. Pupils’ mean score on “Perception of Competence” in the Post-test survey features widely across the various subsets of data as the best predictor of their scores on PSD items in the Delayed-test.

“Awareness of Others” also appears frequently as a good predictor. On the other hand “Awareness of the Outdoor Environment” and “Perception of Confidence” are not good predictors.

There are differences between boys and girls, but not as clear as they were in the relationship between Pre-test and Post-test. For boys, the best predictor has changed from “Attitude towards Groups” between Pre-test and Post-test to “Perception of Competence” and “Awareness of Others” between Post-test and Delayed-test. For girls, the best predictor has remained “Perception of Competence” throughout. These differences might be interpreted to show that for girls, personal and social competence is central to their PSD skills, whereas for boys the drivers for their personal and social skills are mixed. Such inferences from the results are, in truth, rather optimistic, given the quality of the data, but nevertheless they are interesting to ponder.

For the lowest-scoring 50% of pupils the pattern of predictors is much the same as it is for the whole dataset. This cohort of pupils retained their Post-test gains in perception of their PSD skills, and we can surmise that it was an improved view of their personal competence that helped them to do it.

### **3.5 Summary of Chapter**

In this section, the many and various initial interpretations and inferences based on the data analysis are drawn together under key questions relevant to this project. In drawing together the results, many of the more speculative details are filtered out. These questions address the personal and social development of school pupils at the late primary stage of schooling.

*What are pupils’ views of their strengths and weaknesses in personal and social skills?*

Using the most basic information of the mean scores in the four PSD factors, it appears that pupils feel most positive about their social competence and their attitude towards groups, and least positive about their confidence. This picture is consistent across all five administrations of the questionnaire in both the Main Study and the Control Experiment and for both boys and girls. Because the Control Experiment yields the same result as the Main Study, this picture seems to be independent of the residential week intervention and independent of the two groups of schools used in the two experiments.

In addition, girls commonly outscore boys. i.e. feel more positive. In the Main Study this is a consistent feature of the data, and the difference between boys and girls scores of about 4% yields a small but statistically-significant effect size. This is not a consistent feature in the Control Experiment, however, where girls score lower than boys in some areas of PSD and there is no statistically significant difference between boys' and girls' mean scores.

Pupils engaged in the Main Study consistently outscore their counterparts in the Control Experiment by just over 4%. When tested on the Pre-test data, before the intervention in the Main Study could distort the picture, a statistically-significant but only very small effect was measured. There are several reasons why the group of schools in the Main Study might generate higher scores: a sense of anticipation due to the impending week of adventure activity, the higher socio-economic status attributed to this group, contents of the school curriculum and school ethos.

*How do pupils' mean scores change from before to after the residential week of OAA? What development have they experienced?*

It is a feature of the data that pupils who score high in one test score high in other tests, and *vice versa*, those that score low in one score low in others. Also, any changes from one test to another are small relative to pupils' absolute scores, e.g. a change of up to 0.1 relative to a mean score of 3.5 (about 3%) is typical.

Heralded by the analysis of scores on items, but very clear in the pupils' mean scores, is a small but statistically-significant increase from the Pre-test to the Post-test of the Main Study. The increase is seen in all areas of personal and social skills and in both boys and girls. In contrast, in the Control Experiment there is no significant increase between Pre-test and Post-test; in fact, girls' scores decrease. It is reasonable to attribute this change to the benefits of the residential week for PSD.

During preliminary considerations for the regression analysis it became necessary to address the possibility that for any one pupil the increase in PSD score from Pre-test to Post-test is capped by the limitations of the Likert scale, used as an interval scale, at its upper end, e.g. a Pre-test mean score of 3.8 can only be improved upon by a maximum of 0.2. Having said that, on average, high-scoring pupils actually decreased their score slightly from Pre-test to

Post-test and the overall increase derives from substantial increases for the low-scoring pupils.

*In subsequent weeks or months, do pupils retain or lose any changes they record around the time of the OAA week?*

In the Main Study, two to three months after the Post-test a Delayed-test administration of the questionnaire shows that scores decreased substantially by over 3% of the Post-test scores. Again, this is seen in all areas of PSD and in boys and girls. It is a statistically-significant change of small to medium size and takes the scoring to a level that is below that of the Pre-test. A possible interpretation of this is that the pupils experienced an anticipation effect around the time of the residential week, which elevated their scores above a normal baseline recorded by the Delayed-test. This does not contradict the possibility of the increase between Pre-test and Post-test being real.

Extending the investigation of low-scoring and high-scoring pupils to the Delayed-test, in which scores decreased, shows that while high-scoring pupils registered a substantial decrease, low-scoring pupils retained the gains they made at the Post-test. If the results can be believed, this retention of benefit amongst low scorers is one of the most important outcomes of the study.

*Is it possible to predict changes in pupils' PSD from before to after the residential week?*

Much of social science research is aimed at understanding and predicting the effect of interventions on society. This project follows this theme by considering whether a pupil's perception of their personal and social skills can be used to predict the benefit, or development, they might gain from a residential week of OAA. We have just seen that pupils who do not have a very high perception of their personal and social skills seem to benefit more than those that do, and they retain that benefit.

A factor on Awareness of the Outdoor Environment was included in the questionnaire with the aim of discovering whether a pupil's level of awareness influences the change that he/she perceives. In both univariate and multivariate analyses of the data it is clear that it has little or no influence. It is thought that at the late primary stage of schooling, "the outdoors", and even the residential outdoor centre visited by the pupils, is seen as simply an extension of the school playground or back garden rather than a particular kind of environment, *viz.* natural rather than urban. It is interesting, and perhaps relevant, that a survey for the Learning and

Teaching Scotland Outdoor Connections programme found that pupils have difficulty in expressing their emotions around their relationship with nature and the environment (Nicol et al., 2007).

Univariate regressions of Main Study data show that, taken on their own, scores on Pre-test PSD factors are all more-or-less equally good at predicting pupils Post-test perception of their PSD skills, although they hint at some differences between boys and girls. Using multivariate regression, the differences between boys and girls emerge in almost diametrically opposite directions. For boys, Attitude towards Groups is clearly the best predictor; for girls, Perception of Competence is the best, but Awareness of Others and Perception of Confidence are also important. When the multivariate regression analysis is extended to the Delayed-test, to discover what Post-test scores might predict Delayed-test scores, there are significant changes in this pattern in that Attitude towards Groups is rejected as a predictor for boys and Perception of Confidence is rejected as a predictor for girls. For both genders Perception of Competence is an important predictor. At this point, inferences from the results become tenuous, other than to say that this result is a further illustration of differences between boys and girls.

## **Chapter 4                    Implications for research methodology**

### **4.1        Introduction**

The following two chapters discuss the implications of the results of the data analyses in Chapter 3, integrating knowledge gleaned from the development and testing of the questionnaire as described in Chapter 2, and organising them into implications for research methodology and implications for practitioners. The implications for methodology arise from the dichotomy discussed in Chapter 1 between wholly positive qualitative measures and rather mixed quantitative measures of the benefit that OAA delivers for the PSD of children. They extend to a number of effects observed during data analysis, such as a possible euphoria effect, that are, on the whole, undesirable and introduce ambiguity into any conclusions from the study. The implications for practitioners are themselves organised into those for teachers, those for outdoor centres and those for policy makers and funders. The last of these brings us back to the overarching aim of this project, which was to generate evidence to support the cross-party call in the Scottish Parliament for a compulsory residential experience of outdoor adventure activities for all school children (Allardyce, 2008; BBC, 2008).

### **4.2        Conflicting research outcomes**

The research methodology chosen for this study was a classic quantitative one, making use of a questionnaire to measure the change in pupil's perception of their own personal and social skills before and after an OAA intervention. This choice of method was motivated, in part, by the emergence of a worrying distinction in the research literature between the wholly positive anecdotal and qualitative evidence for the benefits of OAA for PSD and the rather mixed results from quantitative research. We saw in the Introduction that even within one study it is possible to obtain conflicting evidence for the magnitude of the benefit, e.g. in Christie's (2004) mixed methods study. In the meta-analysis of Hattie et al. (1997) many negative effect sizes were found despite almost universal belief that residential OAA is beneficial for PSD. It is interesting to consider this conflict in the context of the statement from Greig et al. (2007,p.6) that, "At a "grass roots" level it does not matter very much if one method or another is used as long as the process is rigorous and systematic, and ... based upon sound, reliable and valid data ..". This leads to the question of whether the method used does actually matter. In theory it shouldn't, but it would if there are flaws or limitations

in one or more methodologies. In the following sections, possible limitations of the quantitative methodology are discussed, focusing on the role and nature of the variables. However, we have evidence from this project that there is also a certain lack of rigour arising from the design of the questionnaire, and therefore from the quality of the data, and that will also be explored in the discussion.

#### **4.2.1 The importance of variables<sup>1</sup>**

In the Introduction it was noted that there is a large number of variables that might markedly affect the impact of a residential week of outdoor adventurous activities on the PSD of participants. This is particularly noticeable in the meta-analyses of Cason and Gillis (1994) and Hattie et al. (1997), which identify 11 and 13 variables respectively, and more could be identified that they did not use, such as embedding the experience in the curriculum with preparatory and follow-up work. In Rickinson et al. (2004) 14 “factors” that might influence outcome are discussed in their section 7.2. Sibthorp (2003) discusses 11 variables plus the multidimensional measure of self-efficacy. In the study reported in this thesis the variables were the five factors in the questionnaire, gender, school group, and pupils divided into low-scorers and high-scorers. Assumed to be constant were the stage of development of the pupils, duration of the residential, the programme of activities, the excellence of instruction and embedding the experience in the curriculum. However, unless all potential variables are constrained equally well in all research projects, i.e. genuinely constant or measured reliably and accurately on a suitably large sample, it is more than likely that the research community will not arrive at a consistent set of quantitative measures of benefit. To some extent, this limitation of the method is addressed by using meta-analyses, which are consequently particularly valuable in this field of research. Unfortunately, following the publications of three meta-analyses in the 1990s, there was a shift in research emphasis in this field of education away from a quantitative approach towards a qualitative approach, almost as though researchers felt that they had done as much as they could to cope with the many variables. Insufficient quantitative research has been conducted since then to make another meta-analysis worthwhile and has led to repeated calls for rigorous empirical research (e.g. Rickinson et al., 2004). This project has been a small contribution to that call.

---

<sup>1</sup> In this discussion, “constants” are simply fixed variables: what is chosen to be constant in one study might be a variable in another study.



#### 4.2.2 Consequences of the measurement of variables

The fact that qualitative and anecdotal evidence consistently suggests greater PSD benefits from OAA than quantitative measures do raises the question of whether poorly-constrained variables have the potential to force anomalously low measures of benefit. This might be explained using the equation for effect size, the commonly published measure of benefit,

$$d = \frac{m_A - m_B}{s} \quad (\text{Cohen, 1988}),$$

in which the difference in the mean scores of two tests is divided by a standard deviation from the data, sometimes from the later test, sometimes the mean  $s$ . The consequence of having poorly constrained variables in the study, or constants that are not really constants, will probably be to increase the standard deviations of the mean scores and reduce the effect size. It is possible, therefore, that it is the presence of so many, often poorly-constrained, variables and consequent large standard deviations in studies of the impact of OAA that causes the discrepancy between qualitative and quantitative results. Indeed, Cohen himself says, putting the argument the other way round, “the control of various sources of variation through the use of improved research design serves to increase effect sizes” (Cohen, op.cit, p.13). On the other hand, in their meta-analysis of effect sizes, Cason and Gillis (1993) made a point of investigating the way effect size varied with quality of research design and found an inverse relationship, quite contradictory to Cohen’s theory. Although their definition of research design was not explicit about the way variables were controlled, it is a surprising result that should be investigated further.

This consequence can be illustrated with the data from the Main Study, which yielded the following descriptives:

**Table 4.1** Descriptive statistics for pupils’ mean scores in the Main Study.

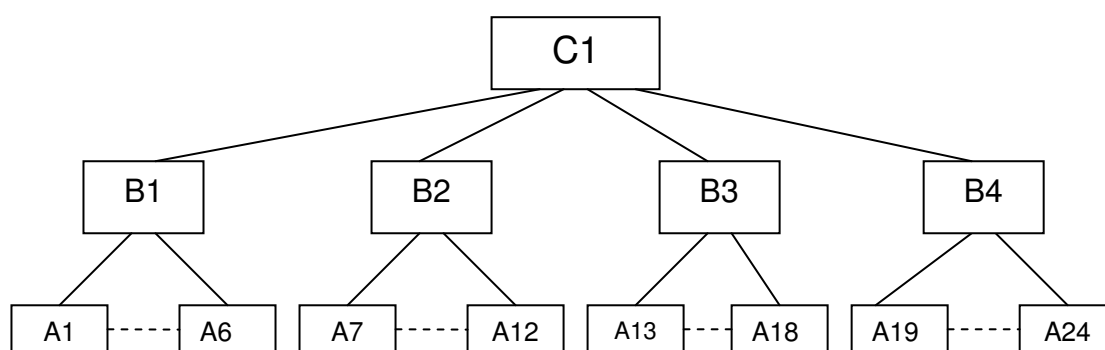
	<b>Pre-test mean</b>	<b>Pre-test standard deviation</b>	<b>Post-test mean</b>	<b>Post-test standard deviation</b>	<b>Delayed-test mean</b>	<b>Delayed-test standard deviation</b>
Boys	3.32	0.38	3.38	0.42	3.24	0.47
Girls	3.44	0.33	3.51	0.36	3.39	0.38

We see that between the Pre-test and the Post-test the mean score for boys increased by 0.06 and for girls by 0.07, an advantage to girls of about 16%, but when couched in terms of effect size using the equation above (with an average standard deviation) the values are 0.15

for boys and 0.20 for girls, an advantage to girls of about 33%. Furthermore, when the scores decline from the Post-test to the Delayed-test, in terms of mean scores boys decline more, 0.14 as opposed to 0.12 for girls, but in terms of effect size, girls decline more, 0.32 as opposed to 0.31 for boys. In both examples, the larger standard deviations in the boys scores act to decrease their effect size, we might say for the worse in the case of Pre-test to Post-test but for the better from Post-test to Delayed-test.

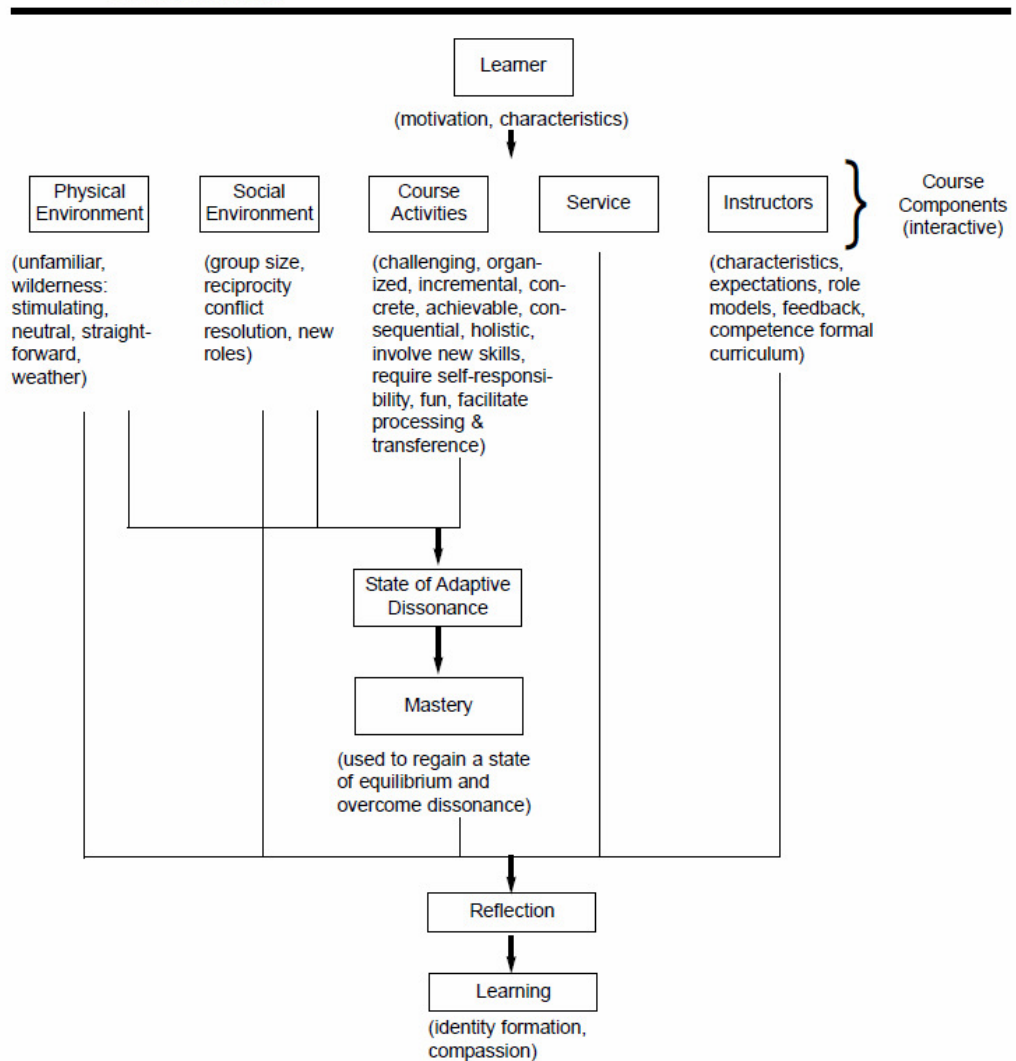
So, it is essential that errors on mean scores, i.e., standard deviations, are constrained as well as they can be. Most of the variables used in studies of personal and social skills are latent variables, i.e., cannot be directly measured, and will, therefore, have an error associated with them. In the case of this research, the variables used for modelling and interpretations were created at various hierarchical levels, as illustrated schematically in Figure 4.1 for PSD: at Level C the one variable was PSD, at Level B the four variables were the PSD factors, and at Level A were the 24 items in the questionnaire. Since all analysis was undertaken on this data set alone, items at Level A could be averaged at Level B or at Level C, so that errors were not compounded as we worked up the hierarchy. However, for example, had we wished to weight pupil's scores on their PSD skills in inverse proportion to the socio-economic background of their school, and that socio-economic background was a latent

**Figure 4.1** The hierarchical structure of variables



**Figure 4.2** A “parallel processing” object-oriented model of student learning during OAA. Figure 2 from McKenzie (2003).

**Figure 2. Building on Walsh and Golins’ (1976) Outward Bound process: An alternative model of student learning.**



variable with an error based on a number of measures of the school, then the error on the PSD score and the error on the school score would compound and, in all probability, create a larger error (standard deviation) to be used in the equation for effect size. The possibility of compounding errors arises in a relatively new model of the student learning process during outdoor adventure activities (McKenzie, 2003) (Figure 4.2). The model has emerged from research with participants in Outward Bound courses and is different from previous models in highlighting the parallel processes that must go on. There are two levels at which a number of different types of measurement are brought together in a latent variable – “state of

adaptive dissonance” and “reflection”. Care would have to be exercised to make sure that errors are not being compounded at these levels.

Finally, a word about constants. If in doubt about the soundness of a constant, it would be wise to measure or categorise it instead! This advice should perhaps have been heeded in this project, since some components that were assumed to be constants at the outset were found to be useful variables, namely the two different groups of schools and the performance of low scorers as opposed to high scorers amongst the pupils. Luckily, these two components could be converted from constants to variables in the course of the study.

#### **4.2.3 Defining variables in the questionnaire**

An attempt was made to tailor the quantitative methodology, and in particular the questionnaire, to the nature of the participants by using concepts and language the schoolchildren would be familiar with and questionnaire structure that would be easy to understand. However, the lack of reliability of the questionnaire compared with reliability measures of other, published questionnaires, suggests that this approach has not worked. Comparing the reliability of my questionnaire with the LEQ of Neill (2007) using a measure that they have in common, it is seen that LEQ yields Cronbach alpha values of 0.8-0.9 for 3-item factors whereas my questionnaire registers 0.5-0.7 for 6-item factors. Looking at factor loadings for individual items, in my questionnaire they are typically 0.5-0.8 whereas in LEQ and in the instrument designed by Tashakkori and Kennedy (1993), which was used in modified form by Nundy (1999), they are 0.7-0.9.

Greig et al. (2007) say that, “Questionnaires can be a good way of finding out about children’s attitudes, and they can be designed to cover exactly the areas you are interested in.” Although I feel confident the contents and structure of the new questionnaire were suitable for the participants I was working with, the nature of the questions was too broad, generated scatter in the responses and consequently reduced the reliability. The number of facets of, for example, self-perception of confidence, addressed by the six questions on that topic contrasts strongly with published, heavily tested and widely used questionnaires. My belief was that there are a number of areas of a child’s life in which confidence plays a major role, and that questions addressing a number of these would yield a good, average score. My questionnaire included four factors of six questions addressing personal and social skills, but might have been better to have eight factors of three questions, each of the three asking

essentially the same thing. A good example of such an approach is provided by the three questions on self confidence in LEQ-H:

*“I know I have the ability to do anything I want to do”*  
*“When I apply myself to something I am confident I will succeed”*  
*“I believe I can do it”*

It is not surprising that these three questions will generate strong factor loadings.

On the other hand, and on a positive note, the answers to the individual questions in the factors in my questionnaire yielded interesting insights into children’s attitudes that would not be obvious from the three questions in the LEQ. Again taking the factor of self confidence as an example, question 18 (Q18):

*“I look forward to presenting my work to my class”,*

received low scores relative to other questions throughout the Main Study as follows.

**Table 4.2** Main Study mean scores on all 30 questions, six “Self-perception of confidence” questions and question 18.

		All participants	Boys	Girls
All 30 questions	Pre-test	3.38	3.32	3.44
	Post-test	3.45	3.38	3.51
	Delayed-test	3.31	3.24	3.39
Confidence questions	Pre-test	3.18	3.18	3.19
	Post-test	3.29	3.28	3.30
	Delayed-test	3.13	3.09	3.18
Q18	Pre-test	2.83	2.72	2.94
	Post-test	2.89	2.81	2.97
	Delayed-test	2.69	2.58	2.81

We see here some familiar trends, such as girls scoring higher than boys and the marked decrease in scores in the Delayed-test, but we now see that on Q18 scores are typically 10-15% lower than the mean score on the confidence questions. This is not to say that scores on Q18 do not correlate with the other scores in the confidence factor; in fact, correlations across the factor are uniform, indicating that pupils who score relatively high do so on all six questions and those that score relatively low do so on all six questions. Although scores on the confidence factor are lower than scores on the whole questionnaire, it is particularly clear that, within this trend, late primary school children are not particularly confident about presenting their work to their class. This information might not have emerged from a more focused questionnaire.

One other clear instance of an outstanding score is on Q6, “I find nature programmes on TV interesting” from the factor “Awareness of the outdoor environment”. This question yields by far the lowest mean score in the whole questionnaire. The outdoor environment factor comes closest of all the factors to consistently breaking down into two components, one component including questions 1 & 11 and the other 6, 16, 21 & 26. A quick look at Appendix 2.1 will show that 1 & 11 are couched in terms of “outdoors” and the other questions in terms of “countryside”, “nature” and “environment”. It was a mistake to use such overly diverse terminology, but nevertheless interesting that the pupils were able to distinguish between the generic “outdoors” and a group of more specific terms. In the context of this distinction, it is clear that the television is not a medium that they find attractive for learning about natural history. Again, the diversity of questions has (by accident) highlighted a difference that would be worthy of further study in due course, which might not have been detected by a more focussed questionnaire.

Certainly questionnaires comprising factors that are very tightly defined by their question content work well. In this study the poor reliability of the questionnaire, as measured by median Spearman’s rho coefficients (Garson, 2009), has reduced the certainty of the outcomes, although the pattern of results is consistent from administration to administration of the questionnaire and probably contains some truth. The enrichment of this study by using factors that are less tightly defined by their question content makes a case for designing questionnaires that contain “the best of both worlds”.

#### **4.2.4 Capping effects on a Likert scale**

There is a potential impact on the magnitude of effect sizes by the capping effect of a Likert scale, specifically when used as an interval scale to measure change from before to after an intervention and particularly when scores cluster towards one end of the scale or the other and become skewed in their distribution, as was the case here. This phenomenon, as it affected this study, is described in section 3.4.1, where it was shown that it might have limited the positive change in scores between the Pre-test and Post-test of the Main Study caused by a clustering of scores towards the top end of the 4-point scale. It was pointed out that if a pupil registered a mean score of, for example, 3.8 in the Pre-test, the maximum benefit the pupil could record as a result of the intervention was 0.2; at the same time, a pupil who registered only 2.5 in the Pre-test could record a benefit of up to 1.5. This effect might have influenced the trend illustrated in Figure 3.6c), which was interpreted to show that low-

scoring pupils benefited more from the intervention than high-scoring pupils, although in that case the capping effect was judged not to be so large as to have had a great impact. The questions were posed, “Is the capping effect a flaw in the use of a Likert scale when data sets plot towards one end of the scale or the other? Is skewness in such data sets an inevitability?” These questions are discussed here.

There is a large volume of literature on the construction and use of Likert scales, including discussion of inherent advantages and disadvantages, but in none of this have I found reference to a capping or limiting effect. This is perhaps because, strictly speaking, a Likert scale is an ordinal scale that should be interpreted using the modal scores, rather than the mean scores derived here, in which case any capping effect and skewness are unlikely to be relevant. However, Loewenthal (2001), and many others, point out that a Likert scale can also be used as an interval scale, with Loewenthal going on to say, “Likert’s suggestion was that the intervals in such rating scales should be *equal-appearing*”. This was the approach taken in this project, which permitted the calculation and extensive use of a mean score for pupils. It is in this situation - using a Likert scale as an interval rather than ordinal scale - that a capping effect becomes possible; but it is only if scores plot towards one end of the scale or the other that skewness in their distribution is likely. With regard to a solution to this problem, a longer scale, such as 8 points rather than the 4 used here, might encourage respondents to avoid the extremes. Finding words to describe the 8 points on this scale, with the more subtle differences in the meaning of the points that this would require, would be difficult, however, especially for a scale constructed for use with children. Indeed, construction of the Likert scale used here was informed by the research of Borgers et al. (2004), who concluded that for good questionnaire reliability, “.. offering about four response options is optimal with children as respondents.” Many other researchers agree that a short scale of simple, clear-cut choices is best for children. However, as we have seen, the short scale, as it was used in the context of this project, was vulnerable to a capping effect and a skewed data distribution. This creates a dilemma for the researcher that must be considered during survey design.

The widespread belief amongst researchers that, in general, short Likert scales are best for children is a good starting point for the construction of a questionnaire scale. However, it is the author’s opinion that the type of survey to be undertaken should be considered before following this guidance closely. In the case of this project, in which children of 10-12 years of age were asked to score their own personal and social skills, and differences in mean score

derived from this, a greater awareness of how the children were likely to respond, which could have been obtained from their teachers beforehand, might have lead to the construction of a slightly longer or differently-worded scale that avoided a capping effect.

### **4.3 Interesting but unwanted effects**

In Chapter 4 we saw that Main Study scores were higher than both Pilot Study and Control Experiment scores. Furthermore, within the Main Study, scores increased from the Pre-test to the Post-test but then decreased markedly in the Delayed-test. A number of possible effects could have been at work to explain these changes. In this section we consider the implications for methodology of the distribution of scores in the Main Study relative to the other experiments.

Also considered briefly is the validity of the assumption that all the schoolchildren who participated in this study were at the same stage in their intellectual and emotional development.

#### **4.3.1 A school group effect?**

The Main Study was carried out with a different group of schools from that used in both the Pilot Study and the Control Experiment. This happened because the schools used in the earlier experiments were not undertaking their normal residential visit at the time it was necessary to collect Main Study data. We have noted that scores are generally higher from the Main Study schools.

A possible interpretation of these results is that there was a school group effect in favour of the schools in the Main Study, perhaps driven by school ethos, but also by socio-economic factors. To investigate the latter, the percentage of pupils eligible for Free School Meals (FSM) in each school can be used as an indicator of the socio-economic environment of the two school groups. If only the unweighted %FSM values of the schools are used as an indicator of socio-economic background, then a Mann-Whitney test for independent samples yields an effect size of 0.54 with a 2-tailed statistical significance of  $p=0.045$  for the difference between the two school groups. Perhaps more meaningful, using the numbers of pupils attending the residential week from each school as a weighting factor, the weighted mean %FSM for the two groups of schools differs by a factor of two: Main Study schools =



13.2%; Pilot and Control schools = 29.9%. This is strong evidence that the pupils in the Main Study came from a better socio-economic environment than those in the Control Experiment as measured by FSM.

Given that there is a strong link between socio-economic background and cognitive development amongst children (Blanden & McNally, 2006), it is possible that the higher test scores in the Main Study reflect a greater awareness of the issues addressed in this project in the Main Study group of schools. Obviously, a schools variable was undesirable from the point of view of comparing results from different experiments and, ideally, should have been avoided, but, on the other hand, it provided an insight into the possible impact of socio-economic background on pupils' self belief.

#### **4.3.2 A euphoria effect?**

The Control Experiment was carried out without the pupils knowing that the survey was associated with their residential week, which was due to take place several months later in that school year. On the other hand, the Main Study was, of necessity, closely linked to the pupils' residential week, the Pre-test taking place the week before and the Post-test taking place the week after. In Chapter 3 the possibility was raised of there being a sense of anticipation amongst the Main Study pupils, which might have inflated their scores at the time of the residential week. The pupils' raised sense of anticipation or excitement relative to its normal level might have created a context variable that had not been anticipated. The term "euphoria", employed here for this effect, is borrowed from Marsh et al. (1986), who used it specifically for its impact on the Post-test, but here I have extended its meaning to include the heightened sense of anticipation immediately before the adventure experience as well.

To try to find other evidence for a euphoria effect in pupils' emotions associated with a week of outdoor adventure a rather cursory analysis was carried out of the exit questionnaire issued to pupils on their final day at Benmore Centre, one of the outdoor centres used in this project. Table 4.3 summarises the findings. The staff at the Centre provided all the questionnaire returns they still had for the last two years, totalling many hundreds. I selected a total of 200 for analysis, focusing on the schools that participated in the Pilot Study and a comparator group, in order to obtain a "feel" for the children's reactions and emotions associated with the week of OAA. First, I looked through the 200 returns to identify 3-5

common responses to each question. There were no differences between the Pilot Study and comparator groups. Then I went through the 200 again to place all responses into one of the 3-5 responses identified for each question. This was a very crude analysis, but it yielded a useful insight into pupils' likely feelings. It is assumed that a similar response would be obtained at the other outdoor centres used in this project. Table 4.3 makes clear the sense of anticipation associated with the Benmore visit, with 42% feeling apprehensive and 51% feeling excited (Question 2). Interestingly, as far as creating an inflated view of personal and social skills is concerned, it can be argued that these two commonly-felt emotions might pull in opposite directions, with apprehension having the potential to reduce scores rather than raise them. Likewise, the feeling of sadness at having to leave the outdoor centre (Question 3) might operate in the opposite direction to a euphoria effect. If this is so, then the higher scores seen in the Main Study, as compared with the Control Experiment, might be more easily interpreted as a genuine reflection of pupils' perception of the personal and social skills or the school effect discussed above.

**Table 4.3** Analysis of exit questionnaire returns from pupils attending a one-week course of adventure activities at Benmore Centre

**Question 1:**

***What do you think of your course at Benmore?***

4 common responses identified:	% of returns
<i>Would like to do more</i>	13
<i>A great, fun experience</i>	56
<i>Some feature of the course was particularly appealing, e.g. instruction, type of activity</i>	23
<i>Mixed feelings about liking all or part of the course</i>	8

**Question 2:**

***Think back to the start of the course. How did you feel then?***

3 common responses identified:	% of returns
<i>Apprehensive about the course</i>	42
<i>Excited by the prospect of the course</i>	51
<i>Doubtful of own ability</i>	7

**Question 3:*****How do you feel now? (With reference back to Q2).***

5 common responses identified:	% of returns
<i>A positive feeling, such as "great"</i>	34
<i>A feeling of greater confidence</i>	7
<i>A sense of achievement, feeling proud or happy</i>	25
<i>A reflective mood, especially a sad feeling at having to leave</i>	27
<i>A feeling of relief that it is all over and can go home</i>	6

**Question 4:*****What have you learnt about yourself this week?***

5 common responses identified:	% of returns
<i>Learnt that determination can take you a long way</i>	16
<i>Realised I have the confidence and/or ability to achieve</i>	36
<i>Realised I can get on with others</i>	9
<i>Learnt I can do specific things or overcome fears</i>	36
<i>Learnt nothing</i>	3

**Question 5:*****What have you learnt about the rest of your group?***

5 common responses identified:	% of returns
<i>A group can be supportive and helpful</i>	38
<i>A group can be sociable</i>	19
<i>Appreciation of the mixed abilities and attitudes of my group, positive and negative</i>	33
<i>Found it difficult to work in a group</i>	7
<i>Discovered I could do "X" by working in a group</i>	4

**Question 6:*****What did you enjoy about the activities?***

3 common responses identified:	% of returns
<i>The challenge they presented</i>	20
<i>The sense of achievement gained</i>	20
<i>They were fun</i>	60

**Question 7:*****Are there any ways in which you would like the course to be different?***

4 common responses identified:	% of returns
<i>Would like more time at the centre to reflect</i>	8
<i>Miss home and family</i>	8
<i>Would like more of a specific activity</i>	69
<i>Would like not to have done a specific activity</i>	16

**Question 8*****Write down the 5 best things about this week.***

	% of things
<i>Adventure activities</i>	78
<i>Non-Adventure activities, e.g. social, housekeeping</i>	22

How might we unravel a euphoria effect causing higher scores in the Main Study Pre- and Post-tests from a school effect described in the previous section? In the Delayed-test the Main Study scores fell to a mean value more in keeping with the scores registered in the Control Experiment. This feature of the data would be more consistent with a euphoria effect around the time of the residential visit than a school effect, which we would expect to exist between all mean scores from the Main Study (Pre, Post and Delayed) and the Control Experiment scores. There is also some evidence from the exit questionnaires that a sense of euphoria exists both prior to (Question 2) and after (Question 3) the adventure week, although some feelings of apprehension and sadness might counter this. On the other hand, there is evidence that the socio-economic background of the two school groups is different and this might translate into mean scores. It is probably fair to say there is some evidence for both effects occurring, but since both are likely to impact on both Pre- and Post-tests, this is not the cause for the increase in Main Study mean scores from Pre-test to Post-test, which is, thus, most likely due to PSD itself.

In terms of implications for research methodology, a euphoria effect would be seen as undesirable and efforts made to remove it or counteract it. The most obvious way of counteracting it would be to remove the Pre- and Post-tests to a time some weeks before and after the experience, perhaps 3 or 4 weeks. A useful review of literature on the subject by Bogner (1998), in his study of the impact of outdoor ecological education on 10-16 year old school children, led him to wait somewhat longer, one month after the intervention, before

carrying out the Post-test. A similar period of time in advance of the intervention for the Pre-test might be sensible. Christie (2004) attempted to use these longer lead and lag times.

### **4.3.3 The importance of children's learning development**

In section 4.1.3 the impact of the rather wide-ranging nature of the questions in the questionnaire on the scatter of pupils' scores was alluded to. However, amongst other causes of scatter in the data one of the most important to this particular study is the varying rate of intellectual and emotional development of the participating pupils. It is clear from a brief review of relevant learning processes, presented in Appendix 4.1, that there is considerable potential for members of a group of 10-12 year old children to be at very different stages of cognitive and, more particularly, affective development, which will influence the pupils' responses to the questionnaire. This is undoubtedly an interesting effect, and might, in addition to introducing more scatter than is desirable, relate to the skewed distribution of the data through the prompting of socially-desirable answers amongst some pupils.

The implication for research is that the 10-12 year old age group is a particularly difficult one to work with from the point of view of ease of measurement of PSD, and yet it is the one in which children commonly have their first experience of residential outdoor learning with the potential for greatest benefit. It is likely that a significant amount of scatter will be introduced into the data simply because the children will be at different stages of development. In this study this scatter ranges from what might be rather diffident responses from some boys (not much more than 2 on the 4-point Likert scale) to the socially-acceptable responses from some girls (4 on the 4-point Likert scale). It is possible that the P6-7 groups that participated in this project should have been restricted to P7 in order for their stage of development and place in the curriculum to be considered as a true constant, as it was assumed to be. On the other hand, it is not clear that this would have made a marked improvement given the rapid and differing rate at which cognitive and affective development is taking place at this age. For research generally, the point can be made that it is important to appreciate not just the components of the OAA experience and the products of learning, but also the processes of learning that will be at work during the experiments, and use this to inform the choice of research methodology, such as questionnaire design. In this project, varying rates of learning development emerged as a potential cause of scatter in the data, therefore, for research generally, it is also important to refer to it when interpreting project results.

## **Chapter 5            Implications for practitioners**

### **5.1      Introduction**

In outdoor adventurous activities and the personal and social development of pupils, the practitioners “at the coal face” are the pupils’ teachers at school and the staff at the outdoor centres, in particular the instructors who look after the pupils on a day-to-day basis during their residential visit. Since the overarching aim of this project was to find quantitative evidence of the benefits of OAA for PSD in order to influence policy makers and funders, this discussion is equally relevant to them. However, given concern over the reliability of the results presented here, it would seem presumptuous to offer any suggestions to practitioners for changing practice. Nevertheless, taking the results at face value some interesting issues arise, the implications of which for practice are now discussed. The results I have obtained are couched in terms of the average pupil and it must be acknowledged that for the teacher and instructor it is each individual pupil that matters, perhaps guided by these generic outcomes.

### **5.2      Implications for teachers**

The summary table of pupils’ exit questionnaire returns from Benmore Centre (Table 4.3) is based on a sample of only 200, but given the widespread anecdotal and semi-qualitative evidence in the research literature that concurs with the findings, we can take it to be representative for late primary school children. The evidence in the Table makes it quite clear that pupils thoroughly enjoy the outdoor adventure experience. About 90% return positive comments about their personal experiences of the outdoor and other activities. Coupled with the small but statistically significant increase in PSD scores from the Pre-test to the Post-test of the Main Study, we can conclude that important benefits are realised. It is particularly pleasing to see low-scoring pupils benefiting quite substantially on the PSD measures. This rewarding experience for the pupils is also a rewarding experience for the teachers and outdoor centre staff.

A key question is whether the short-term personal and social benefits just described are retained in the weeks and months following the residential week. Hanna (1995) found that with a group of adults participating in environmental education, long-term benefits decreased unless participants remained committed to their post-programme goals or they received

mentoring. Hattie et al. (1997) recorded a positive benefit for personal development of effect size 0.17 between post-test and delayed-tests in their meta-analysis. They said “the continued gains and longevity of the follow-up effects are the most impressive findings ... unique in education literature”, but then went on to say “caveats ... about which aspects of the programs are effective need to be heeded”, in other words, they could not say why the long-term gains came about. Rickinson et al. (2004) identified the need for effective follow-up in order to reap the benefits of fieldwork but did not explain where the evidence for this comes from. On the other hand, questionnaire returns from teachers attending Low Bank Ground Outdoor Centre in the Lake District provide evidence of follow-up work related to the curriculum, although in this case without any quantitative measure of the benefit (Cooper, 2004). Moore (1990) goes so far as to say that to gain benefit from outdoor adventure activities they should be fully integrated into the curriculum. Overall, there is both evidence for and opinion in favour of follow-up work on residential experiences in order to retain benefit, and this probably applies in both cognitive and affective learning domains.

In this project, the marked fall in pupils’ mean scores between the Post-test and the Delayed-test of the Main Study indicates a serious loss of short-term benefit, although broken down into low-scoring and high-scoring pupils it appears that the benefits were actually retained by the low-scoring pupils. This last feature is particularly welcome and perhaps illustrates the effectiveness of outdoor activities to facilitate change for the better in less confident young people. It suggests that it is not impossible to retain the short-term benefits, as the meta-analysis of Hattie et al. (op.cit.) has shown. It is also ironic, if there is any link between low-scoring pupils, lack of self belief and lower socio-economic status, that the pupils who seem to benefit most might be those whose parents can least afford to send them on the residential week.

Although the causes of long-term retention of benefit have not been forthcoming in previous research, it would seem that using the residential week to provide material for subsequent curriculum studies is a real possibility. Indeed, the outdoor learning component of A Curriculum for Excellence (Learning and Teaching Scotland, 2010) promotes just this sort of approach – take an outdoor activity and map onto it other curriculum components. Essay or poem writing, environmental and social studies, health and wellbeing, mechanics, communication, and PSD itself, are all examples of curriculum areas that could be enriched by the children’s experiences of outdoor adventurous activities. As it happens, through the

use of these other curriculum areas there is a means of measuring benefit in more traditional, academic ways. Personal and social experiences during the residential week have the potential to provide good material for the areas of literature and health and wellbeing. In the past, how PSD is measured has been a cause for concern (Inman *et al.*, 1998), and this concern might have contributed to the reluctance to support OAA with hard-pressed education funds.

In the process of working with the schools in this project, the class teachers for the pupils undertaking a residential week were asked to complete a short questionnaire aimed at discovering how much use is made of the experience across the curriculum. The results were mixed. Only eight questionnaires were returned, but these were supported by several conversations with teachers on this subject. The results from the questionnaires were as follows:

- Four indicated that the experience is integrated into the curriculum and four indicated that it isn't
- Four indicated that reference is made to the experience for the whole or a substantial part of the year whereas four said reference is made only around the time of the visit (not the same groups of four as above)
- The main reason for not making more use of the experience was the fact that not every member of the class had been on the visit
- The areas of the curriculum making use of the experience were Health and Wellbeing (includes PSD) (5 returns), Writing/Literacy/Expressive Arts (5 returns), Geography/Social Science/Fieldwork (3), PE (2), ICT (1), All! (1).

This rather patchy follow-up could be the reason the short-term benefits are not converted into long-term benefits. Considering the substantial cost and effort that goes into providing a residential week for pupils it would seem sensible to reap the maximum benefit from it.

With an explicit outdoor learning component in A Curriculum for Excellence this situation might change, including training or CPD for teachers who do not yet understand the benefits of outdoor education (Nicol *et al.*, 2007). However, until every child is able to participate in the residential experience it is unlikely schools will wholeheartedly integrate it into the curriculum. It is worth noting, however, that the one school that does integrate the experience into all parts of the curriculum throughout the year has the second highest incidence of free school meals and several pupils who could not participate. It is important to know how the pupils who could not participate were included in the follow-up work, but



assuming that this is done sensitively, this example shows that with the motivation from the teachers the maximum use of the experience can be achieved. Partly because the sample from this school is so small (17) and partly because I did not want to analyse data by individual schools, whether this one school has achieved greater benefit is not known, but it would be a useful exercise in further research.

This issue contains a strong message for policy makers and funders. Maximum benefit from a residential week of OAA is gained when the experience is integrated into work in the classroom, both before and in the weeks and months subsequent to the residential visit, and this work can contribute significantly to the objectives of A Curriculum for Excellence. However, the maximum benefits are not being realised because work in the classroom would exclude pupils whose parents cannot afford to sponsor the residential week. Ironically, this research seems to show that the pupils who would benefit most are likely to be those whose parents can least afford to send them on the residential week

Other results of interest to teachers are those on gender differences, the trends in performance of low scorers, and the irrelevance of awareness of the outdoor environment on outcome. These results might influence the approach a teacher takes during preparation and follow-up for the residential week. Further comments on the relevance of these to practitioners are included in the next section.

### **5.3 Implications for outdoor centres**

It was not the intention of this project to examine the quantity or quality of provision of residential weeks of OAA at outdoor centres, even though there is a thread in the research literature that focuses on quality as a variable, and in particular on the philosophy and quality of instruction. On the other hand, the role of the outdoor centre is central to the overarching aim of this research, which is to support the case for full funding for all school children to attend a residential week of adventure activities at such a centre.

During this project I have gained the impression that, as things stand at the moment, demand from schools and supply from outdoor centres are well matched in this area of education, although this balance has been reached without any national policy on provision behind it (Nicol *et al.*, 2007). The school teachers I met felt that the centres were doing an excellent job, and were it not for limitations from cash and staff resources on their side, demand would

increase. The last two years at primary school present an ideal time for children to undertake a residential week of OAA, and the mix of activities on offer from the outdoor centres fits the bill perfectly in terms of personal and social development. As pointed out in the previous section, there is also potential for using the experience to enhance the curriculum.

Nevertheless, the situation is a precarious one. I was aware of cancellations from schools who had booked well in advance only to find that parental contributions were not forthcoming as the time of the visit approached – schools are reluctant to undertake the visit if only half the class can attend, which would hinder integration of the experience into the curriculum and, hence, maximising the benefits. On the supply side, all centres in Scotland now have to pay their way on at least a recurrent basis, although local authority centres can apply for grants for capital expenditure (Andy Beveridge, personal communication). Any reduction in the large school customer base, estimated at 300,000 pupil-days per year by Higgins (2002), would create serious financial problems for LEA centres in particular, since they meet the bulk of this demand.

Thus, currently, the balance between supply and demand is a fine one. Sector provision for schools is only likely to grow if education funding is found to permit more schools to participate. In his article for the Sunday Times, Allardyce quotes a cost of £10m to permit 53,000 pupils to attend an outdoor centre for one week (Allardyce, 2008). This is a sum of just under £200 per head - £200 is just under 5% of the annual budget for each primary school pupil. The 53,000 he quotes is about one seventh of the primary school role in Scotland and presumably represents the year group (usually P6 or P7) that will undertake the visit. In a time of austerity, new money will be very difficult to obtain, no matter how worthy the cause, but redirection of expenditure, in light of the outdoor learning component of A Curriculum for Excellence, is a greater possibility. It would seem sensible, therefore, for outdoor centres to consider how they might diversify their course offerings to better match the needs in more academic areas.

The development of a reliable, year-on-year measure of the outcomes of OAA programmes for schoolchildren would, no doubt, be helpful to outdoor centres. In the current tight financial climate, [positive] outcome measures would help to make the case for continuing their work. Since virtually all the qualitative and anecdotal evidence points to considerable benefit for PSD, and most, but not all, quantitative studies point to benefit, the argument is already strong for their continued contribution to the education and development of young people. Even this study, with its rather poor questionnaire, was able to demonstrate short-

term benefits. The closest approach in the published literature to a routine, practical measure of the benefits gained by young people is a questionnaire proposed by Cooper (2004) and developed at Low Bank Ground Outdoor Centre in the Lake District on behalf of Wigan Local Education Authority. Most of the Centre's work is with 9-11 year old schoolchildren undertaking a five-day outdoor adventure programme. The measure is based on information collected from leaders before their visits (the Course Objectives Form), evaluations from leaders (the Course Evaluation Form) and young people (the Evaluation Sheet and Group Statement) during their visit and follow-up questionnaires (the Course Review Form) about one month later. The last of these includes questions about follow-up work in the classroom, as discussed in the section on Implications for Teachers. The quantitative nature of the measure is achieved by asking teachers to score their responses to questions on a scale of 1 (low) to 5 (high). Table 5.1 gives mean scores from the first 60 questionnaire returns, focusing on course objectives.

**Table 5.1** Means of teachers' scores on Low Bank Ground post-course questionnaires (N=60)

<b>Monitoring Pupil Progress: a list of educational objectives</b>	<b>How important do you consider this objective? 1. Not at all → 5. Very.</b>	<b>How far were these objectives met? 1. Not at all → 5. Completely.</b>
To build self-confidence and self-esteem	4.9	4.6
To encourage self-reliance and personal responsibility	4.9	4.5
To develop teamwork and cooperation skills	4.9	4.6
To develop good relationships and respect differences between people	4.8	4.5
To develop problem-solving and decision-making skills	4.3	4.1
To increase choices for a healthy and safer lifestyle	4.1	3.8
To encourage students to take an active role as citizens	4.1	3.8
To encourage environmental awareness and understanding	4.6	4.4
To promote spiritual development	3.8	3.5
To achieve particular curriculum objectives in Geography, English, etc.	3.7	4.4

Although there is an interesting tendency in these results for less important objectives to be less well met, it worth noting an exception to this in "particular curriculum objectives",

which are least important but quite well met. This observation is offered in support of the suggestion above that outdoor centres might like to consider diversifying their course offerings, since they appear to have the potential to meet objectives in curriculum areas.

It is likely that most outdoor centres request feedback and evaluation from their customers. However, this may not be quantitative feedback; and yet, quantitative measures speak very loudly when it comes to making arguments for financial support. I would urge all outdoor centres, not just the ones of interest here that provide residential weeks for schoolchildren, to find a means of including quantitative evaluations in the feedback they request from their customers, and to collect this on an ongoing basis. Policy makers ought to require this kind of measure for outdoor centres used by schools.

Specific results from this project that will be of interest to outdoor centres are those on gender differences, the trends in performance of low scorers, the importance of varying rates of learning development amongst late primary children and the irrelevance of awareness of the outdoor environment on benefit. All of these are outcomes from the study that might influence the approach taken by instructors during adventure activities. In summary:

- Girls tend to score higher than boys on questions associated with their personal and social skills. Perception of Competence and Awareness of Others stand out as the areas of PSD in which girls out-score boys most emphatically. On the other hand, both boys and girls return relatively low scores on their Perception of Confidence. A few boys return quite low scores throughout, contributing to negative skewness in the distribution of pupils' mean scores. It is clear that some boys are not at all confident of their personal and social skills. The intervention did not affect these gender differences. In a regression analysis of what prior PSD skills best predict boys and girls self assessment following the residential week, it was found that for boys the best predictor is Attitude towards Groups whereas for girls it is Perception of Competence followed by Awareness of Others and Perception of Confidence.
- Staying with measures of pupils' awareness of their personal and social skills, it was interesting to discover that the low-scoring pupils appeared to benefit more from their residential experience. Since boys predominate amongst the low scorers, it is predominantly boys who benefit more. Following on from this, in the Delayed-test of the Main Study, low-scorers appear to retain their benefits whereas high scorers lose

them. With a visiting primary school group it is unlikely that instructors will know the identity of low-scoring pupils, but in a broader context, this result is in keeping with the belief that OAA is particularly effective at yielding PSD benefits for young people with low self esteem.

- It would probably not surprise providers and instructors that there is evidence from this project, in the form of the high level of scatter in the children's scores, that the stage of learning development of late primary school participants varies substantially. The theory of learning processes presented in Appendix 4.1 accounts for this variability. Programme design for the residential week will have evolved over many years of seeing what works and what doesn't work in terms of providing desired outcomes for the children and their schools. However, this does not preclude occasionally revisiting the theories and principles of learning processes that lie behind a successful programme. A rich experiential environment and the opportunity for the transfer of learning between cognitive and affective domains as well as between programme components, as described in Appendix 4.1, would exploit the children's learning processes to the full. Although the details of this last statement might be new, the general idea of using the theory of learning processes to inform programme design is not new, and has been investigated by researchers on numerous occasions (e.g. Walsh and Golins (1976), Hopkins and Putnam (1993), McKenzie (2000)).
- During this study an attempt was made to find out if pupils' awareness of the outdoor environment prior to attending the residential week would influence the PSD benefit they experienced. The belief was that a prior awareness would amplify their benefits. In the event, there were limitations to measuring benefit as such, and an investigation was carried out to find out if prior awareness influenced Post-test score. It was found that outdoor awareness did not predict Post-test PSD score at all, indicating that a pupil with little or no awareness had the potential to score highly on measures of PSD just as much as an environmentally-aware pupil might have. In this context, it is interesting to note that in their review of the impact of OAA, Rickinson et al. (2004) found that, "the evidence for a positive link between outdoor adventure activities and environmental understanding and values is not strong". At the outset we noted that the outdoors or natural environment provides one of the plethora of new and different situations that might stimulate young people to learn about themselves, but it does seem as though familiarity with it is not a prerequisite for learning and benefiting.

## 5.4 Implications for policy makers

In the two preceding sections of this Chapter some points have been made that are particularly relevant to the debate that is currently continuing amongst politicians with regard to the provision of residential outdoor education for every Scottish school pupil (TheyWorkForYou, 2010). In summary, the points are these:

- If there is any link between low-scoring pupils, lack of self belief and lower socio-economic status, then we can say that the pupils who seem to benefit most might be those whose parents can least afford to send them on a residential week of outdoor learning, such as a week of OAA.
- There are obvious links to be made between the activities on a residential week of outdoor learning and the objectives of the outdoor learning component of A Curriculum for Excellence (Learning and Teaching Scotland, 2010), in this case, in the PSD area.
- Moreover, maximum benefit from a residential week of OAA is gained when the experience is integrated into work in the classroom. However, the maximum benefits are not being realised because teachers are reluctant to undertake work in the classroom that would exclude those pupils whose parents cannot afford to send their child on the residential week.
- Capacity at outdoor centres is only likely to grow if education funding is found to permit more schools to participate in OAA free of charge. It is estimated that £10m per annum would allow all pupils to attend an outdoor centre for one week. This equates to just under £200 per head, just under 5% of the annual budget for each primary school pupil.
- It would seem sensible for outdoor centres to consider how they might diversify their course offerings to better match the needs of more academic areas and of A Curriculum for Excellence. Also, they should collect quantitative evaluations from their customers on an ongoing basis to support their case for funding for every school pupil. Indeed, policy makers ought to require this kind of measure from the outdoor centres used by schools.

Although referring to secondary schools rather than primary schools, a 2009 YouGov poll commissioned by the Scottish Conservatives asked 1,183 adults if they agreed or disagreed that every secondary school pupil in Scotland should receive at least one week's free outdoor education a year, such as hill-walking, abseiling, canoeing or camping. It found that 71% agreed or strongly agreed, 13% disagreed, while 14% neither agreed nor disagreed. To provide one week of outdoor education a year of a residential kind is a remarkably ambitious prospect, costing in excess of £140M or nearly 6% of the whole Scottish education budget per year. Nevertheless, the poll shows an encouraging level of public support, although it raises the question of whether it is possible to support free residential outdoor education at both primary and secondary level. The political debate at the moment suggests that the Scottish National Party prefers to spend what money can be afforded at primary school level whereas the Scottish Conservatives say 11-15 years old and the Liberal Party prefers to spend it at secondary level. The Scottish Labour Party, whilst supporting the concept, does not say where they would prefer to spend money. Elsewhere in this thesis the point is made that children at 10-12 years of age are at an optimal stage in their education to benefit in the area of personal and social development, and Mortlock (1978) also makes a strong case for this being an optimal time.

It is clear that there is both public and cross-party political support for well-resourced outdoor education, including outdoor adventurous activities. I believe it is possible to demonstrate that outdoor education leads to a measurable positive benefit in both PSD (the subject of this research) and, even more convincingly, in academic subjects. It is clear from my conversations with teachers that at primary school level the emphasis of the residential week of OAA is fun, PSD and the transition to high school. My own background in fieldwork-based science and my work on promoting sport and the study of science, society and culture to young people leads me to believe that at the secondary school level hard-pressed outdoor education resources would be best spent in support of academic subjects, with a view to facilitating learning and making key academic subjects more attractive. And yet, despite this overwhelming evidence and opinion in favour of free residential outdoor education for all Scottish school children, there are still barriers to overcome. Since few doubt the value of the experience, these barriers must be fully understood and reduced to a level that can be overcome when education expenditure is prioritised.

## **Chapter 6            Summary of project outcomes**

### **6.1      Summary of key research results**

The overarching aim of this project was to gather some quantitative evidence in support of the cross-party call in the Scottish Parliament for a compulsory residential experience of outdoor adventure activities for all school children, in the belief that this greatly benefits their personal and social development (Allardyce, 2008; BBC, 2008). Personal and social development of the individual is a strong thread running through A Curriculum for Excellence (LTS, 2008). Full funding for 53000 pupils each year would be £10M. At present, less than 20,000 of the total school roll in Scotland of about 700,000 attend a residential outdoor centre, largely at a cost to their families.

This project focused on the potential benefits for children of late primary school age. The experience of primary schools who attend residential outdoor centres with their P6 or P7 pupils has been one of very positive benefits at this stage. Qualitative research has confirmed this, but more relevant to this project is the fact that there is no quantitative evidence from Scotland in support, although there is some quantitative support from elsewhere in the UK and overseas. Moreover, quantitative evidence from elsewhere is not wholly positive. This lack of quantitative evidence is a particular problem at a time when expenditure is allocated against measurable outcomes.

Against this background, the five objectives of this project were:

- To design and test a new questionnaire that could be used to make quantitative measures of P6/7 pupils' perception of their own personal and social skills,
- To carry out an experiment to make measures just before, just after and about 2-3 months after the pupils' residential experience of OAA at an outdoor centre,
- To analyse the data to identify changes in scores that might indicate benefit, and test these for statistical significance,



- To further analyse the data to discover whether there were certain attributes amongst children of late primary age – environmental awareness, personal and social awareness, etc. – that make them particularly amenable to benefit, and
- To consider the implications of the results for researchers and practitioners, including policy-makers and funders of residential visits, and understand the implications in the context of theories and models of learning that might apply during OAA.

### **6.1.1 Summary of results from Chapter 2**

With the aim of making a quantitative measure of the impact of a residential week of OAA on the PSD of children of 10-12 years old, a new questionnaire was researched and designed specifically for use with children of this age. It was more generic than other questionnaires in the nature of its questions, with the intention of obtaining a better overall picture of the pupils' views. The questions were grouped into four PSD factors and a factor called "Awareness of the Outdoor Environment". The latter was to discover if a greater awareness in this area increased the benefit of OAA. The questionnaire was used with an essentially random sample of pupils, initially to test and refine it (c.100 pupils), then in a Control Experiment (c.100) and, finally in the Main Study (c.350). Some important and interesting results emerged from evaluating the questionnaire.

- Reverse-meaning questions in the initial version of the questionnaire yielded inconsistent results and were subsequently not used.
- Questionnaire validity was tested with pupils, teachers, outdoor centre staff and research colleagues, leading to the selection of 30 items from the original 52.
- Factorisation of the questionnaire was generally poor, but improved in the Main Study and the original five factors were retained in the final 30-item version.
- Both internal and external reliability were generally poor but consistent in the way the various parts of the questionnaire behaved from administration to administration, i.e. some factors consistently performed better than other factors. Interestingly, mean scores on items correlate strongly from Pre-test to Post-test in all parts of the

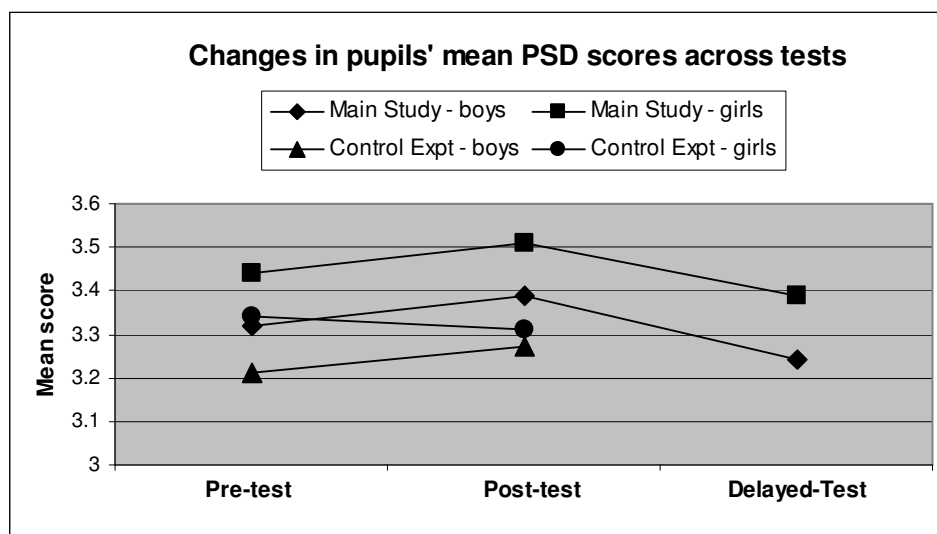
project ( $R^2 \geq 0.9$ ) but, of course, this does not necessarily indicate that pupils are reliably interpreting the questions: this strong correlation coupled with the poor reliability suggests that on any one question some pupils moved their score up and others down from one test to another.

- Most interesting from the point of view of the personal and social skills of the pupils was a persistent signal of increasing internal reliability from Pre-tests to Post-tests, even in the Control Experiment where there was no intervention, and independent of the general level of pupils' mean scores. This was interpreted as a result of pupil familiarity with the questionnaire at the second administration, not so much with individual questions as with the general personal and social concepts, which they learn about in class.

### 6.1.2 Summary of results from Chapter 3

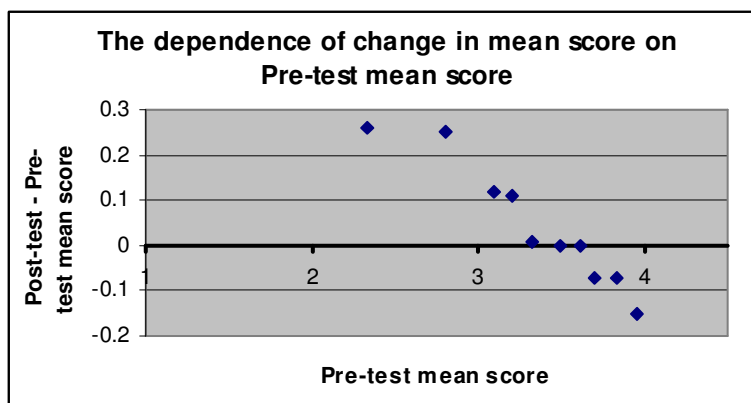
Chapter 3 addressed the analytical objectives of the project. Following an inspection of scores on items in the questionnaire, the changes in pupils mean scores across the three tests of the Main Study were investigated: the Pre-test and Post-test, which were two weeks apart, either side of the OAA intervention, and the Delayed-test, 2-3 months later. Comparison was made with the results of the Control Experiment. The figures that follow illustrate the key results for PSD factors. Beneath each figure is a summary of the results it illustrates and a summary of their possible interpretations.

**Figure 6.1** Changes in pupils' mean PSD scores across tests



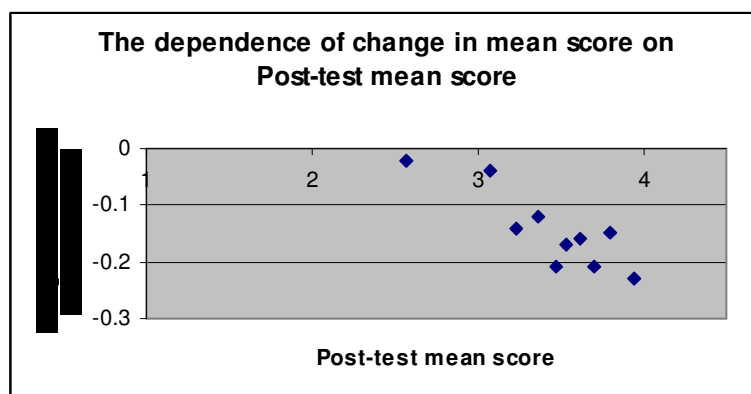
- Throughout the Main Study girls scored higher than boys in a very consistent way. This could reflect greater personal and social skills amongst girls, but it could also reflect a tendency for girls to give socially-acceptable, higher scores. Some boys returned outlying low scores. Interestingly, this gender difference is not as clear in the Control Experiment, which was conducted with a different group of schools that has a lower socio-economic measure on the basis of percentage of free school meals.
- In the Pre-test and Post-test, scores in the Main Study were higher than in the Control Experiment. This might also relate to the use of different school groups with different socio-economic measures, but it might also reflect an anticipation or euphoria effect with the group attending the outdoor centre. For the control group the survey was not associated with any particular event.
- Whereas the previous two results might be associated with “undesirable” effects, the increase in pupils’ mean score from Pre-test to Post-test in the Main Study, for both boys and girls, is likely to be “desirable”. The Control Experiment shows a slight increase for boys, which was not significant, but the girls score actually decreased. This signal is interpreted as an improvement in self perception of personal and social skills resulting from the intervention in the Main Study.
- .
- The fall in scores in the Delayed-test is substantial and statistically significant. It indicates a loss of benefit that could be related to the participating schools being rather patchy in their reinforcement of the experience. It is interesting to see that the Delayed-test score is less than the Pre-test score and more-or-less the same as the scores in the Control Experiment, lending credence to the possibility of a euphoria effect in the Pre- and Post-test of the Main Study.

**Figure 6.2** The dependence of change in score between Pre-test and Post-test on Pre-test mean score for the Main Study, all cases, all PSD items. Graph created in Excel using one-tenth bands of data.



- The whole Main Study sample of 368 was divided into tenths for further study. This graph shows that between the Pre-test and the Post-test the mean PSD score of low-scoring pupils increased by as much as 10% whereas the mean score for high-scoring pupils recorded a fall, although this fall might be influenced to some extent by the capping effect of the 4-point Likert scale. The increase of 10% in self perception of personal and social skills is in keeping with the view that outdoor adventure is therapeutic for less socially-aware children.

**Figure 6.3** The dependence of change in score between Post-test and Delayed-test on Post-test mean score for the Main Study, all cases, all PSD items. Graph created in Excel using one-tenth bands of data.

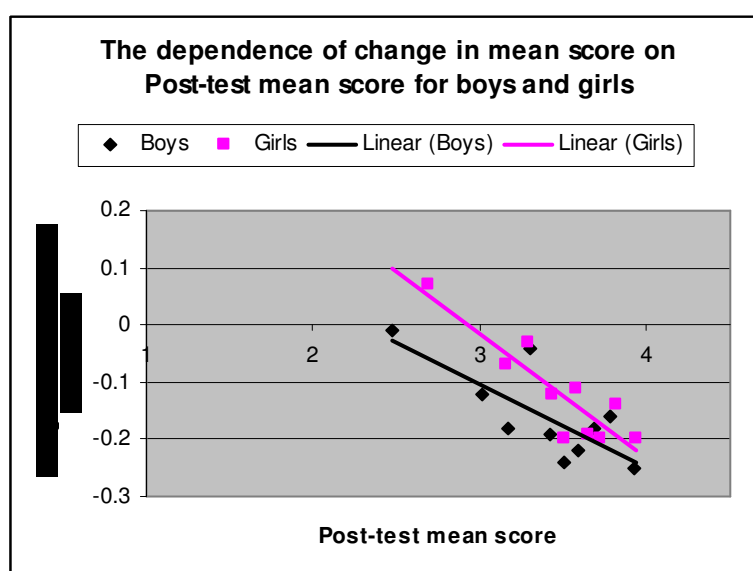


- Following on from the previous observations at the tenths level, this graph shows that between the Post-test and the Delayed-test low-scoring pupils retained their benefit from the intervention with a fall of up to only 0.05 or 1.7%. On the other hand, high-scoring

pupils show a fall of up to 0.2 from Post-test to Delayed-test. Given that the participating schools were rather patchy in their reinforcement of the experience this is a remarkable achievement by the low-scorers, and it is suggested that it is a continuation of the therapeutic effect described in the previous bullet point. Note that as scores decrease there is no capping effect from the Likert scale.

- Here, the results from the previous graph are considered by gender. It has been noted that girls score higher than boys and that boys register some outlying low scores. However, the trend in better performance of low-scorers over high-scorers throughout the Main Study applies to both girls and boys.

**Figure 6.4** As for Figure 6.3 but broken down by gender. Note that each data point now represents a sample of only c.15.



Chapter 3 also described univariate and multivariate regression analyses to discover if it is possible to predict changes in pupils' PSD from before to after the residential week? For example, we have just seen that pupils who do not have a very high perception of their personal and social skills appear to benefit more than those that do, and they seem to retain that benefit. On the other hand, pupils' awareness of the outdoor environment appears to have no influence over the amount of any benefit they might gain in PSD, a result that is broadly consistent with other studies for this age group.

Univariate and multivariate regressions of Main Study data yielded some differences between boys and girls as to what personal and social skill might be the best predictor of their Post-test PSD score following a residential week of OAA. The results are tentative, but for boys it seemed to be their attitude towards group work, whereas for girls it seemed to be their perception of their competence. The loss of benefit in the longitudinal study correlated most strongly with perception of competence for both boys and girls. No further interpretation was made of these results.

## **6.2 Summary of implications**

### **6.2.1 Summary from Chapter 4**

Chapter 4 presented a discussion of the implications of material in Chapters 1 to 3 for some aspects of research methodology. It had not been intended to have an investigation of research methodology as an objective of this project, but a number of unanticipated results from the design of a new questionnaire and its use with 10-12 year old schoolchildren prompted a consideration of the implications of these results for research methods. The implications are chiefly concerned with the use of variables and with undesirable effects on scores that have caused a good deal of ambiguity in the interpretations of the results. The relevant points can be summarised as follows.

- There is a worrying distinction between the wholly positive anecdotal and qualitative evidence for the positive benefits of OAA and the mixed results from quantitative research. Even within one study it is possible to obtain conflicting evidence for benefit. It was suggested that the very large number of potential variables inherent in the provision of outdoor adventure activities and in personal and social development works against the acquisition of consistent quantitative measures of benefit. It tends to force quantitative estimates downwards and, potentially, generate large errors in latent variables. The questionnaire used in this project was not very well designed and introduced a good deal of scatter into the measurements of PSD. The implication for research is that variables must be constrained as tightly as possible: assumed constants must be truly constant and errors on variables as small as possible. If anything, larger sample sizes than those indicated by Cohen's (1982) tables are desirable.

- A number of interesting but undesirable effects on methodology emerged during the analysis of survey data. First, by choosing to work with an age range of 10-12 years old school pupils, narrow in some respects but wide in terms of child development, which is rapid at that age, a relatively large scatter in scores was introduced. It is important that models of child development and learning processes are taken into account when designing a research project so that the project is appropriately and tightly defined. A probable school group effect was introduced by having to work with two groups of schools with different socio-economic backgrounds to meet the project schedule. This effect was difficult to unravel from a possible euphoria effect on the school group that experienced the OAA intervention, in contrast to the control school group that did not.

## **6.2.2 Summary from Chapter 5**

A discussion of implications for practitioners in OAA and PSD was presented in Chapter 5. It was addressed to the pupils' teachers at school and the staff at the outdoor centres, but was also relevant to policy makers and funders. Suggestions arising from the discussion should be treated with caution in view of the reliability of the questionnaire. Nevertheless, taking the results at face value some interesting issues arise.

For teachers, a key result was the undoubted enjoyment and apparently measurable benefit for their pupils of the residential week of OAA. However, the small but significant gains between the Pre-test and the Post-test were more than lost over the 2-3 months between the Post-test and the Delayed-test, and this was attributed to the rather patchy nature of any follow-up or reinforcement work back at school. A silver lining to this cloud was the observation that low-scoring pupils do seem to retain their benefits, but, at the same time, it is also ironic that the pupils who seem to benefit most might be those whose parents can least afford to send them on the residential week.

Outdoor centres are in a precarious funding situation. It would help their case for continued funding if they kept a reliable, quantitative, ongoing measure of the benefits they deliver, such as the measure at Wigan LEA outdoor centres based on quantified pupil and teacher feedback. As far as the delivery of OAA to pupils is concerned, features of the results of this project that might be of interest are the gender differences, the trends in performance of low scorers mentioned above, and the irrelevance of awareness of the outdoor environment amongst pupils on their ability to benefit from their experiences.

This research raises a number of points of detail that will be of interest to policy makers, but the overarching issue is that despite strong public and cross-party political support for well-resourced outdoor education, including outdoor adventure activities, there are still barriers to overcome to achieve sufficient funding for all school children to enjoy a residential week free of charge. These barriers must be fully understood and reduced to a level that can be overcome when education expenditure is prioritised.



## References

- ALLARDYCE, J. (2008, May 25). Outdoor education mandatory for young Scots. *The Sunday Times*.
- BARRETT, J., & GREENAWAY, R. (1995). *Why Adventure?*. Coventry, England: Foundation for Outdoor Adventure.
- BBC. (2008). *Bid to boost outdoor learning*, from <http://news.bbc.co.uk/1/hi/scotland/7411997.stm>
- BLANDEN, J., & MCNALLY, S. (2006). *Poverty 123: Mind the gap: child poverty and educational attainment* from <http://www.cpag.org.uk/info/Povertyarticles/Poverty123/gap.htm>
- BLOOM, B. S. (Ed). (1956). *Taxonomy of Educational Objectives: 1. Cognitive Domain*. London: Longman.
- BLOOM, B. S., KRATHWOHL, D. R., & MASIA, B. B. (1964). *Taxonomy of Educational Objectives: 2. Affective Domain*. New York: David McKay Company, Inc.
- BOGNER, F. X. (1998). The Influence of Short-Term Outdoor Ecology Education on Long-Term Variables of Environmental Perspective. *The Journal of Environmental Education*, 29(4), 17-29.
- BORGERS, N., HOX, J., & SIKKEL, D. (2004). Response Effects in Surveys on Children and Adolescents: The Effects of Number of Response Options, Negative Wording and Neutral Mid-Point. *Quality & Quantity*, 38, 17-33.
- CASON, D., & GILLIS, L. (1994). A meta-analysis of outdoor adventure programming with adolescents. *Journal of Experiential Education*, 17(1), 14-47.
- CHRISTIE, E. A. (2004). *"Raising Achievement" in Secondary Schools?: A Study of Outdoor Experiential Learning*. University of Edinburgh.
- COHEN, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale: Lawrence Erlbaum Associates.
- COHEN, J. (1992). A Power Primer. *Psychological Bulletin*, 112, 155-159.
- COOPER, G. (2004). How School Groups Benefit from Outdoor Education. *Horizons*, 25, 10-18.
- DILLON, J., MORRIS, M., O'DONNELL, L., REID, A., RICKINSON, M., & SCOTT, W. (2005). *Engaging and Learning with the Outdoors - The Final Report of the Outside Classroom in a Rural Context Action Research Project*: National Foundation for Educational Research.
- DISMORE, H., & BAILEY, R. (2005). "If only": Outdoor and adventurous activities and generalised academic development. *Journal of Adventure Education and Outdoor Learning*, 5(1), 9-19.
- EWERT, A. (1989). *Outdoor adventure pursuits*. Columbus, Ohio: Publishing Horizons.
- EYSENCK, M. W., & KEANE, M. T. (1990). *Cognitive Psychology*. Hove, East Sussex: Lawrence Erlbaum Associates Ltd.
- FIELD, A. (2009). *Discovering Statistics using SPSS* (3rd ed.). London: SAGE Publications Ltd.
- GARSON, G. D. (2009). *Reliability Analysis: Statnotes from North Carolina University*, from <http://faculty.chass.ncsu.edu/garson/PA765/reliab.htm>

- GILLIS, H. L., & SPEELMAN, E. (2008). Are Challenge (Ropes) Courses an Effective Tool? A Meta-Analysis. *Journal of Experiential Education*, 31, 5-29.
- GREENAWAY, R. (2010). *Experiential Learning articles and critiques of David Kolb's theory*, from <http://reviewing.co.uk/research/experiential.learning.htm>
- GREIG, A., TAYLOR, J., & MCKAY, T. (2007). *Doing Research with Children* (2nd ed.). London: SAGE.
- GROSS, R. D. (1992). *Psychology: The Science of Mind and Behaviour* (2nd ed.). London: Hodder & Stoughton.
- HANNA, G. (1995). Wilderness-related environmental outcomes of adventure and ecology education programming. *Journal of Environmental Education*, 27(1), 21-32.
- HANS, T. (2000). A meta-analysis of the effects of adventure programming on locus of control. *Journal of Contemporary Psychology*, 30(1), 33-60.
- HASKELL, R. E. (2001). *Transfer of Learning: Cognition, Instruction and Reasoning*. London: Academic Press.
- HATTIE, J., MARSH, H., NEILL, J., & RICHARDS, G. (1997). Adventure education and Outward Bound: out-of-class experiences that make a lasting difference. *Review of Educational Research*, 67(1), 43-87.
- HIGGINS, P. (2002). Outdoor Education in Scotland. *Journal of Adventure Education and Outdoor Learning*, 2(2), 149-168.
- HIGGINS, P., & NICOL, R. (2002). *Learning as Adventure: Theory and Practice*: DfES and Connexions Service.
- HOPKINS, D., & PUTNAM, R. (1993). *Personal Growth Through Adventure*. London: David Fulton Ltd.
- INMAN, S., BUCK, M., & BURKE, H. (Eds). (1998). *Assessing Personal and Social Development: Measuring the unmeasurable?* London: Farmer.
- KOLB, D. (1984). *Experiential learning: Experience as the source of learning and development*. Englewood Cliffs, NJ: Prentice Hall.
- LTS. (2010). *Curriculum for Excellence through Outdoor Learning*. Glasgow: Learning and Teaching Scotland.
- LOEWENTHAL, K. M. (2001). *An Introduction to Psychological Tests and Scales* (2nd ed.). Hove, East Sussex: Psychology Press.
- MARSH, H., RICHARDS, G. & BARNES, J. (1986). Multidimensional self-concepts: A long-term follow-up of the effect of participation in an Outward Bound program. *Personality and Social Psychology Bulletin*, 12, 475-492.
- MCKENZIE, M. (2003). Beyond "The Outward Bound Process": Rethinking Student Learning. *The Journal of Experiential Education*, 26(1), 8-23.
- MCKENZIE, M. (2000). How are adventure education program outcomes achieved?: a review of the literature. *Australian Journal of Outdoor Education*, 5(1), 19-28.
- MCLOUGHLIN, C. (1997). The implications of the research literature on learning styles for the design of instructional material. *Australian Journal of Educational Technology*, 15(3), 222-241.
- MCNAUGHTON, A. H. (1966). Piaget's Theory and Primary School Social Studies. *Educational Review*, 19(1), 23-32.

- MOORE, G. (1990). Adventure Activities for School Children. In J. C. Miles & S. Priest (Eds), *Adventure Education*. State College, PA: Venture Publishing Inc.
- MORTLOCK, C. (1978). *Adventure Education* (2nd ed.). Ambleside: Colin Mortlock.
- MYGIND, E. (2009). A comparison of childrens' statements about social relations and teaching in the classroom and in the outdoor environment. *Journal of Adventure Education and Outdoor Learning*, 9(2), 151-169.
- NEILL, J. T. (2007). *Life Effectiveness Questionnaire: a research tool for measuring personal change*, from <http://www.wilderdom.com/leq.html>
- NEILL, J. T., & RICHARDS, G. E. (1998). Does Outdoor Education Really Work? A Summary of Recent Meta-Analyses. *Australian Journal of Outdoor Education*, 3(1), 2-9.
- NICOL, R., HIGGINS, P., & ROSS, H. (2006). *Outdoor Education: the views of providers from different contexts in Scotland: Learning and Teaching Scotland*.
- NICOL, R., HIGGINS, P., ROSS, H., & MANNION, G. (2007). *Outdoor Education in Scotland: A Summary of Recent Research*. Edinburgh: Scottish Natural Heritage & Learning and Teaching Scotland.
- NUNDY, S. (1999). The Fieldwork Effect: The Role and Impact of Fieldwork in the Upper Primary School *International Journal of Geographical and Environmental Education*, 8, 190-198.
- RICKINSON, M., & OTHERS. (2004). *A Review of Research on Outdoor Learning: National Foundation for Educational Research and King's College London*.
- SHEPHARD, R. J. (1997). Curricular Physical Activity and Academic Performance. *Pediatric Exercise Science*, 9, 113-126.
- SIBTHORP, J. (2003). An Empirical Look at Walsh and Golins' Adventure Education Process Model: Relationships between Antecedent Factors, Perceptions of Characteristics of an Adventure Education Experience, and Changes in Self-Efficacy. *Journal of Leisure Research*, 35, 80-106.
- SOED. (1993). *Guidelines on Personal and Social Development 5-14*, from <http://www.ltscotland.org.uk/5to14/guidelines/psdevelopment.asp>
- STEINBERG, L. (2005). Cognitive and affective development in adolescence. *TRENDS in Cognitive Sciences*, 9(2), xx-xx.
- TASHAKKORI, A., & KENNEDY, E. (1993). Measurement of Self Perception in a Multi-Racial Context: psychometric properties of a modified Self Description Questionnaire. *British Journal of Educational Psychology*, 63, 337-348.
- THEYWORKFORYOU. (2010). *Outdoor Education: Scottish Executive Question Time*, from <http://www.theyworkforyou.com/sp/?id=2010-11-25.30917.2#g30918.0>
- TREIMEN, D. J. (2009). *Quantitative Data Analysis: Doing Research to Test Ideas*. Chichester: Jossey Bass/John Wiley.
- WALSH, V., & GOLINS, G. (1976). *The exploration of the Outward Bound process*. Denver, CO: Outward Bound publications.

## Appendix 2.1 The questionnaire



Scottish Centre for PE, Sport and Leisure Studies  
The Moray House School of Education  
The University of Edinburgh  
St Leonard's Land  
Holyrood Road  
Edinburgh EH8 8AQ

### Questionnaire on your week of outdoor activities

My name is Roger Scrutton, and I work at Edinburgh University.

This questionnaire is part of a project I am doing to find out how a week of adventure activities in the outdoors helps young people to get on in life.

Your answers to the following questions would be very helpful to me.

Thank you very much for your help.

-----

Please write your initials here .....

Please say whether you are male or female .....

For each of the following statements, place a tick in the box that applies to you. For example,

1. I help my parents with jobs around the house.	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me ✓
2. I go online on the computer every evening.	No, this is not like me	This is not really like me ✓	This could be like me	Yes, this is definitely like me

### Statements about you

### How do you feel about this statement?

1. I enjoy being outdoors
2. I like to help with jobs around school
3. I am a confident person
4. I am good at making new friends
5. I like team games
6. I find nature programmes on the TV interesting
7. I think it is important to have hobbies that keep you busy
8. I enjoy trying new experiences
9. I am happy to help other people get their jobs done
10. I am good at listening to other people's points of view
11. I look forward to school classes that take us outdoors
12. I enjoy the challenge of a problem that is difficult

No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me

13. I have the ability to finish what I set out to do	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
14. I like listening to my friends' stories	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
15. I would support my team-mates, even if we are losing	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
16. I think it is interesting to explore the countryside	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
17. I like to be active and get on with things.	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
18. I look forward to presenting my work to my class	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
19. I find it interesting that different people have different abilities	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
20. I would be happy to be chosen as the leader of a team	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
21. I believe it is important to protect the outdoor environment	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
22. I feel proud about the tasks I have successfully completed	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
23. I find it easy to make decisions	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
24. I always listen carefully to instructions	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
25. I prefer working in a group to working on my own	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
26. I would like to help to protect the outdoor environment	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me

27. I always try to do my best	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
28. I stay calm when things go wrong	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
29. I find it easy to talk with adults	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me
30. I learn a lot from working in a group	No, this is not like me	This is not really like me	This could be like me	Yes, this is definitely like me

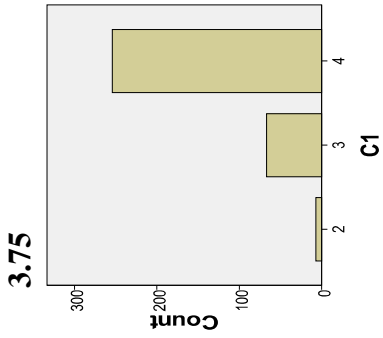
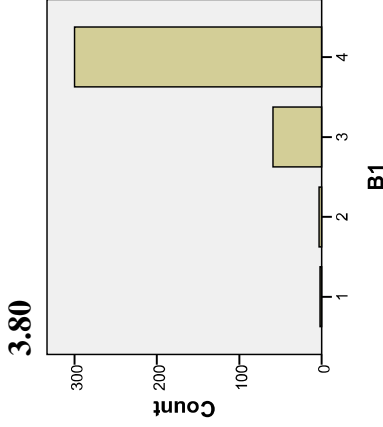
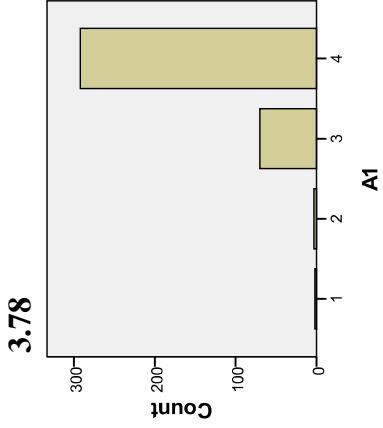
## **Appendix 3.1     Histograms of scores on items**

Histograms are presented for the Main Study and the Control Experiment. In each case the data are arranged in the order of the six questions in each of the five factors. The question is given, and the mean score for the question is top-left of the histogram. For the Main Study there are histograms for the Pre-test (labelled A, 368 cases), Post-test (labelled B, 368 cases) and Delayed-test (labelled C, 328 cases). For the Control Experiment there are histograms for the Pre-test and the Post-test, both 117 cases.

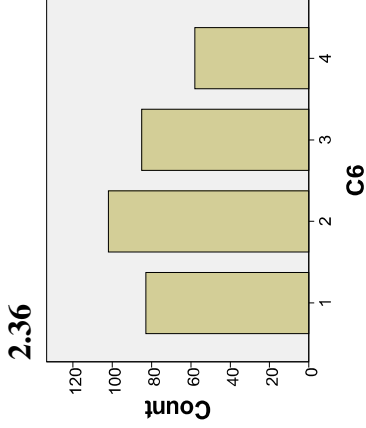
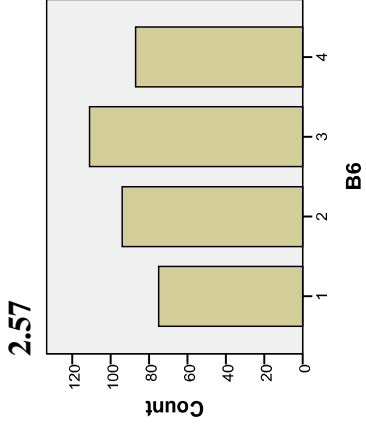
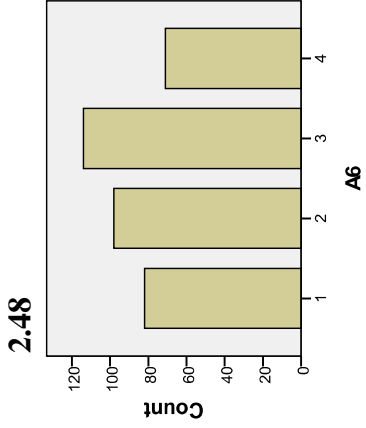


**MAIN STUDY**

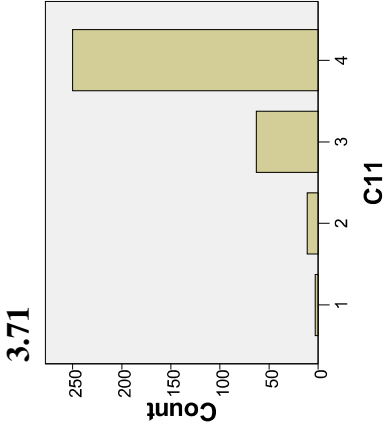
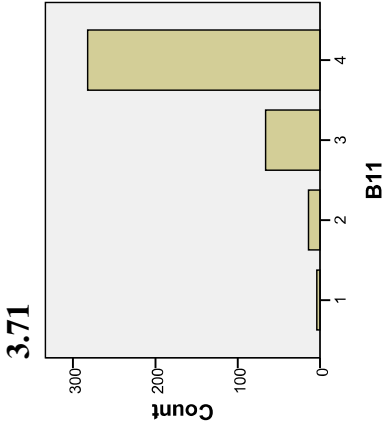
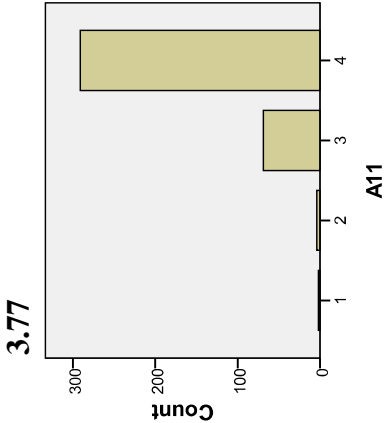
Question 1. I enjoy being outdoors.



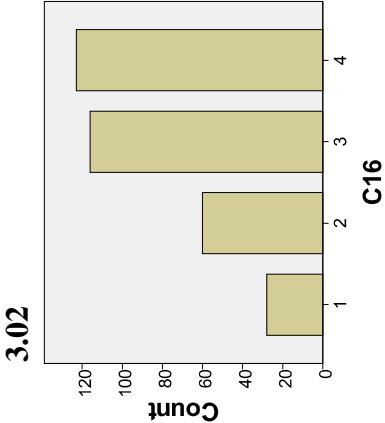
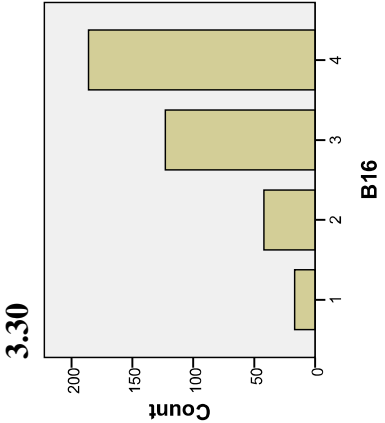
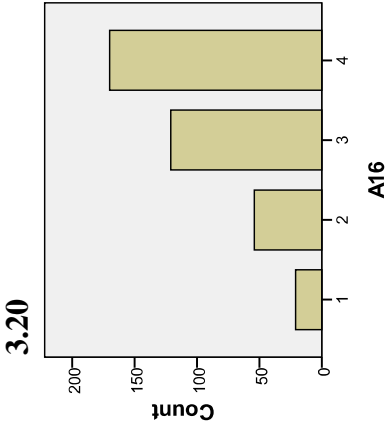
Question 6. I find nature programmes on the TV interesting.



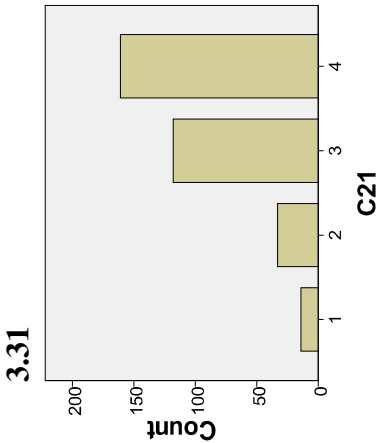
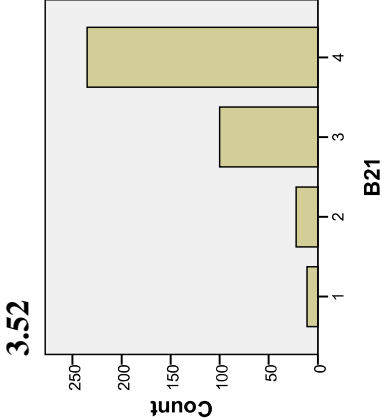
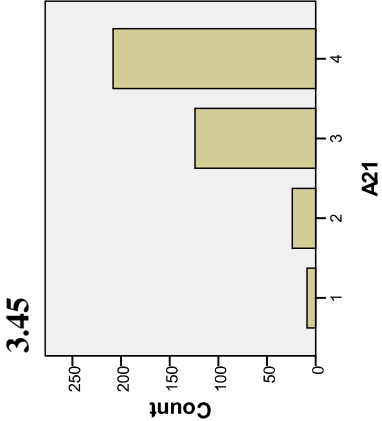
Question 11. I look forward to school classes that take us outdoors.



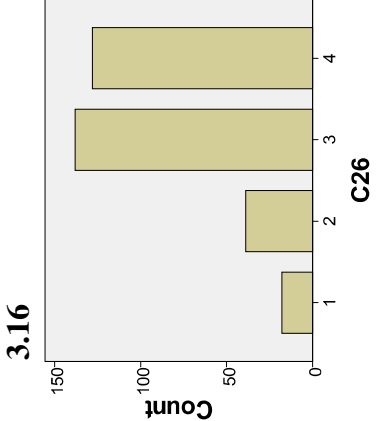
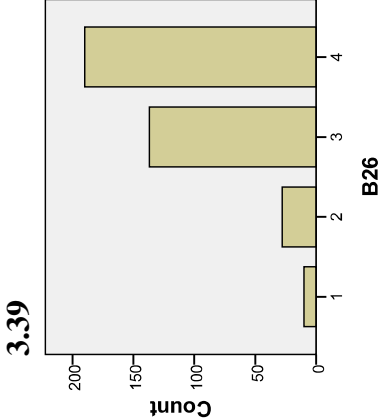
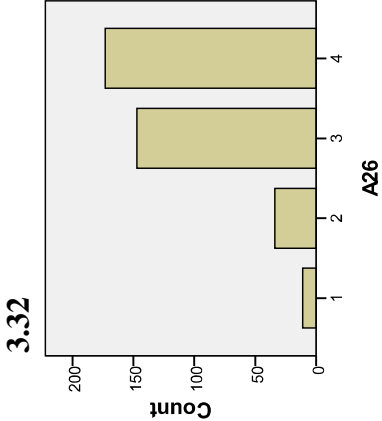
Question 16. I think it is interesting to explore the countryside.



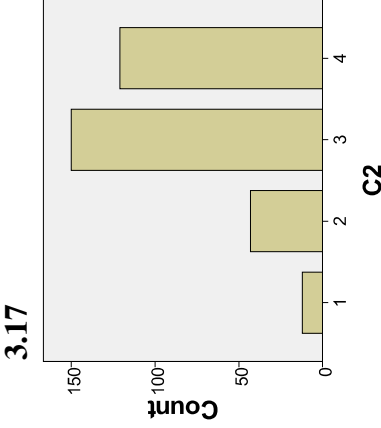
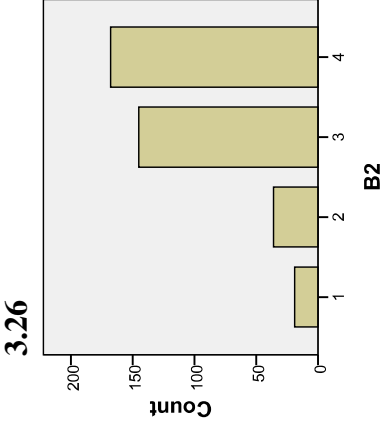
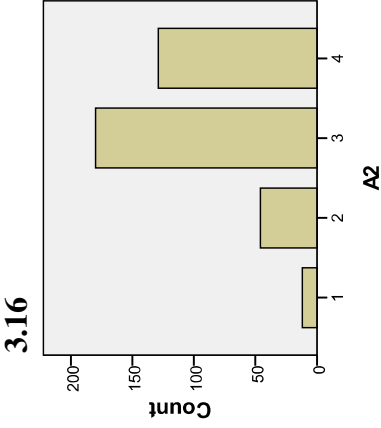
Question 21. I believe it is important to protect the outdoor environment.



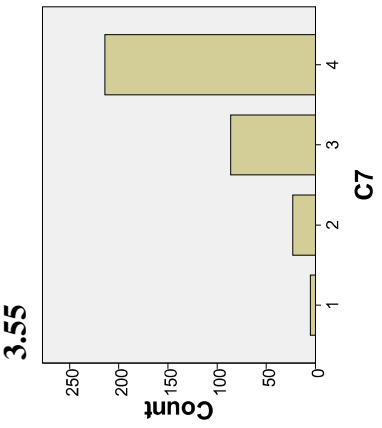
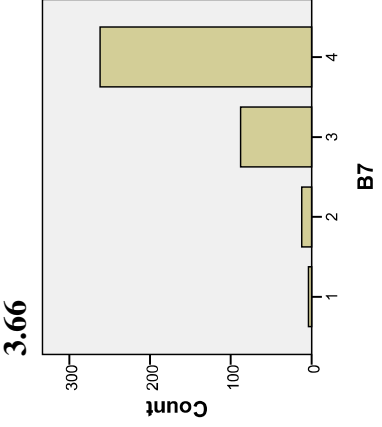
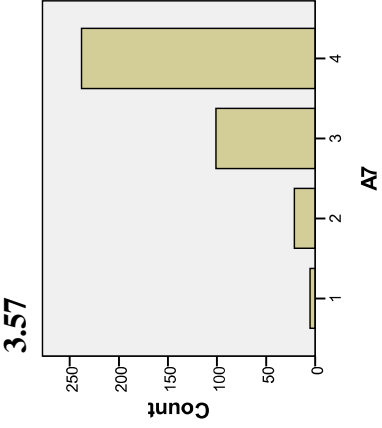
Question 26. I would like to protect the outdoor environment.



Question 2. I like to help with jobs around school.

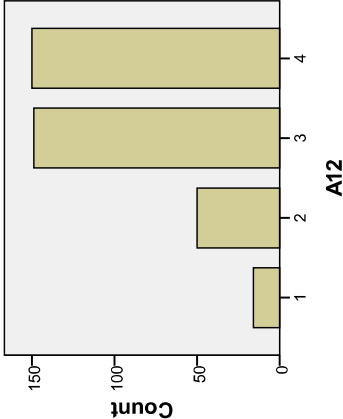


Question 7. I think it is important to have hobbies that keep you busy.

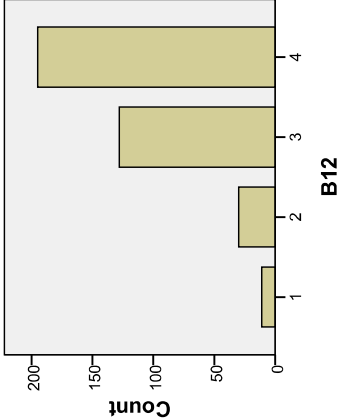


Question 12. I enjoy the challenge of a problem that is difficult.

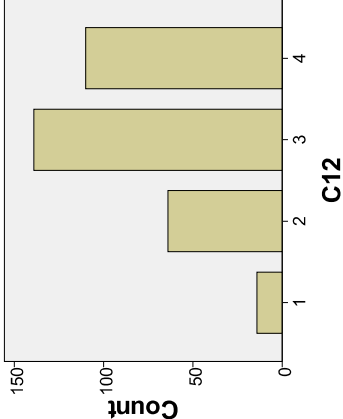
3.19



3.39

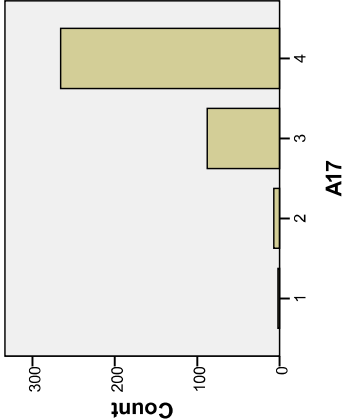


3.06

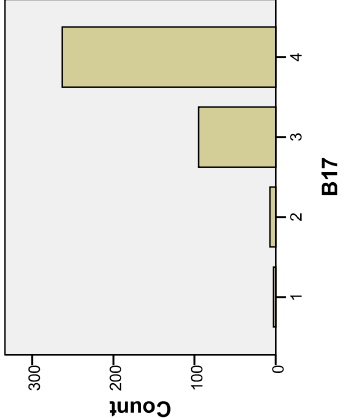


Question 17. I like to be active and get on with things.

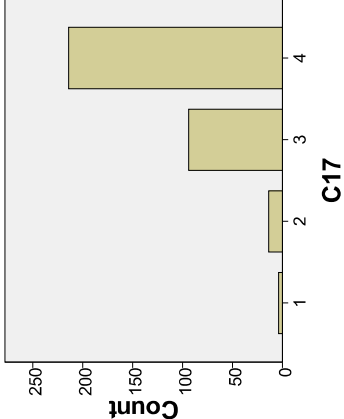
3.70



3.68

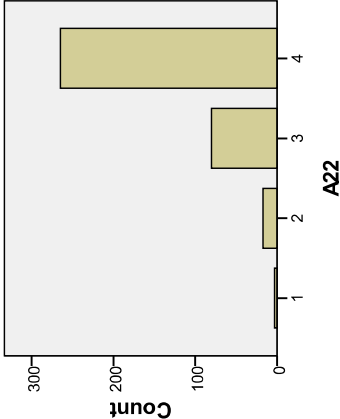


3.59

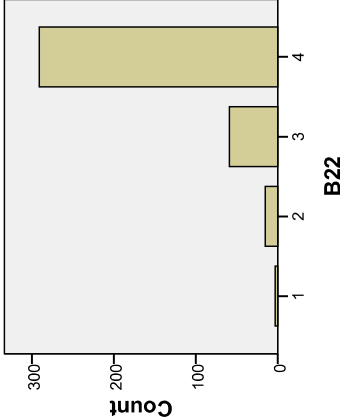


Question 22. I feel proud about the tasks I have successfully completed.

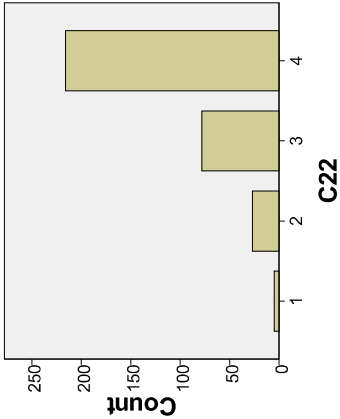
3.66



3.73

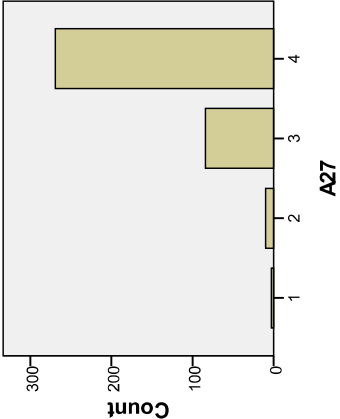


3.55

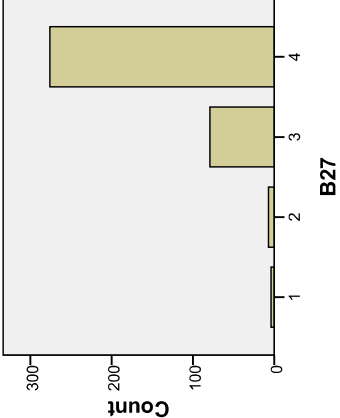


Question 27. I always try to do my best.

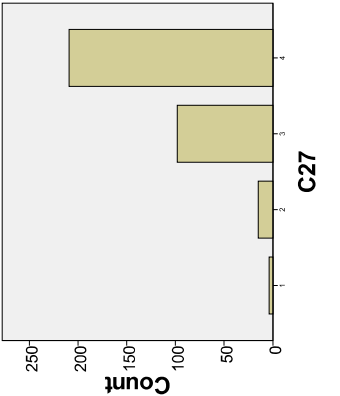
3.69



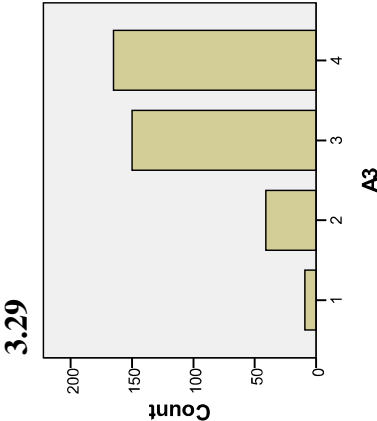
3.71



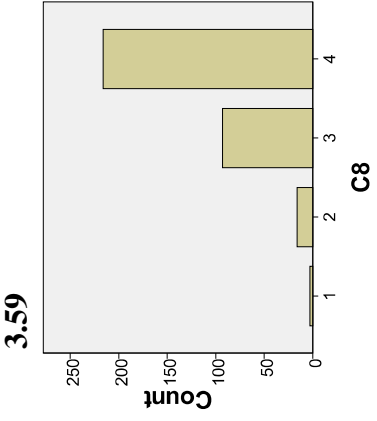
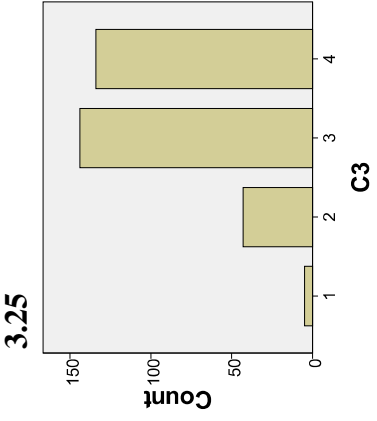
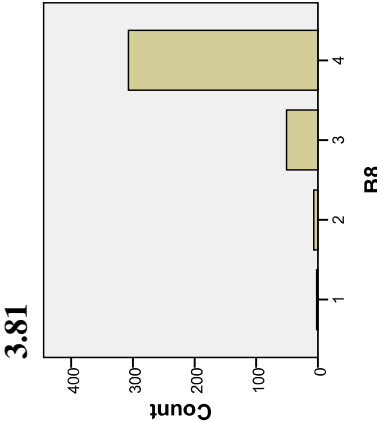
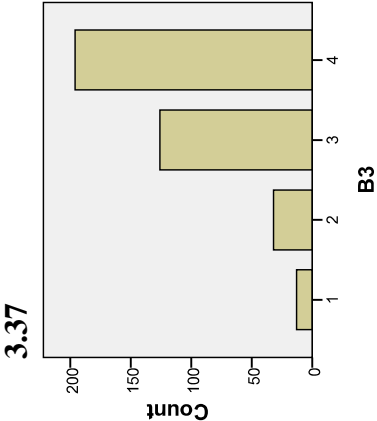
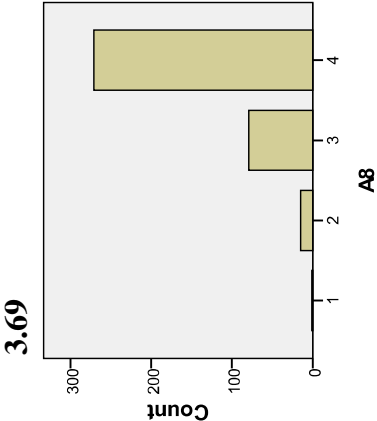
3.57



Question 3. I am a confident person.

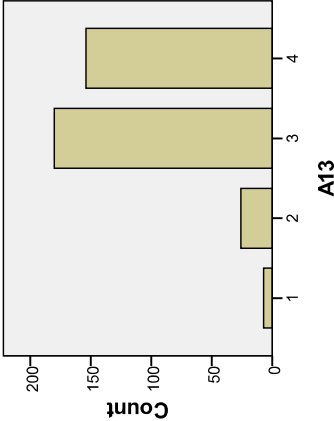


Question 8. I enjoy trying new experiences.

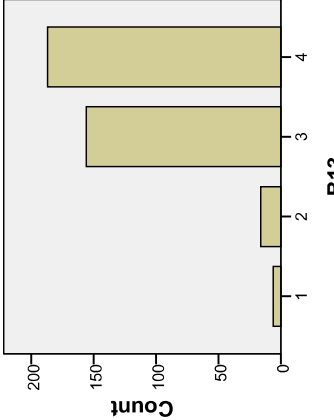


Question 13. I have the ability to finish what I set out to do.

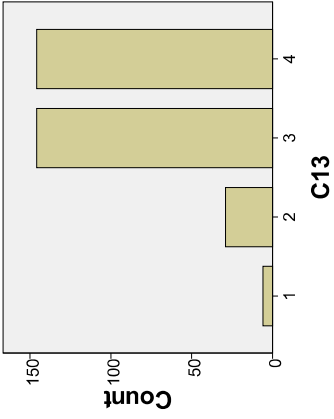
**3.31**



**3.44**

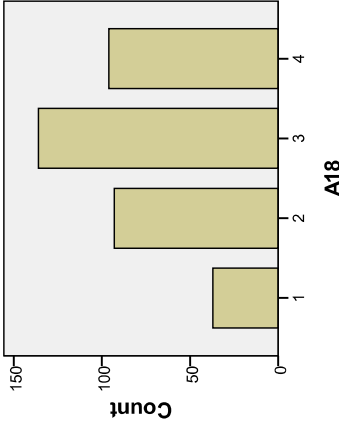


**3.06**

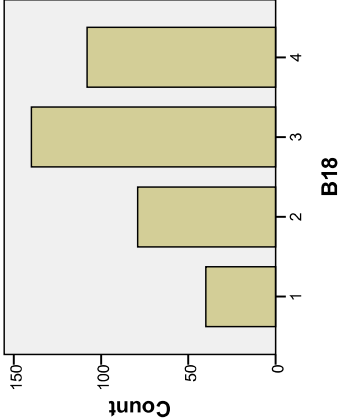


Question 18. I look forward to presenting my work to my class.

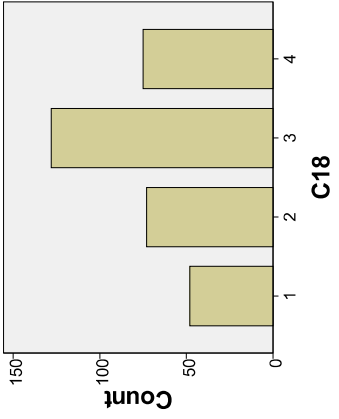
**2.80**



**2.86**

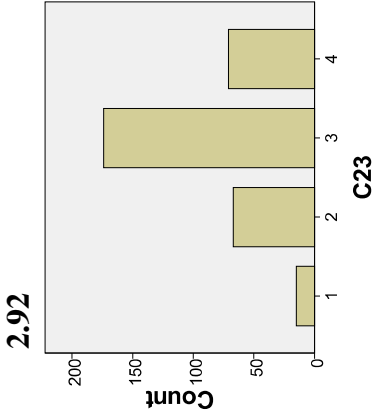
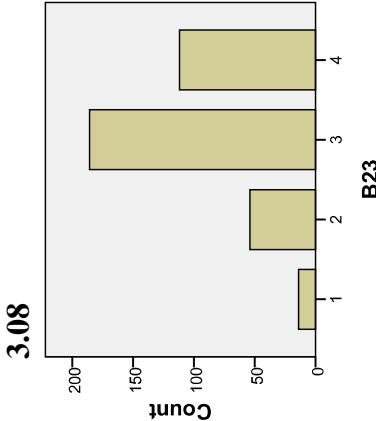
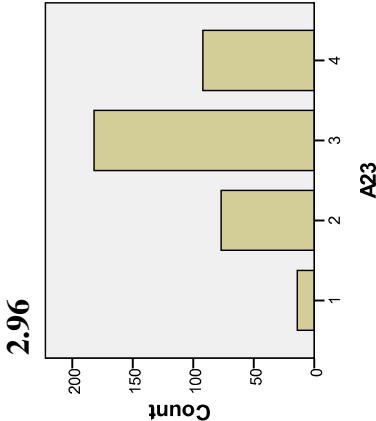


**2.71**

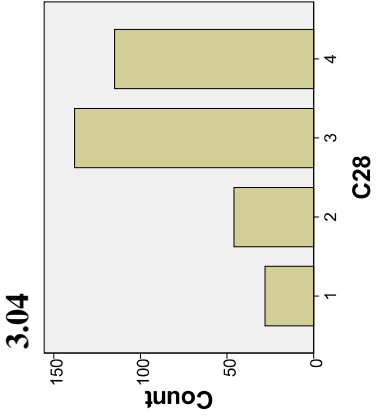
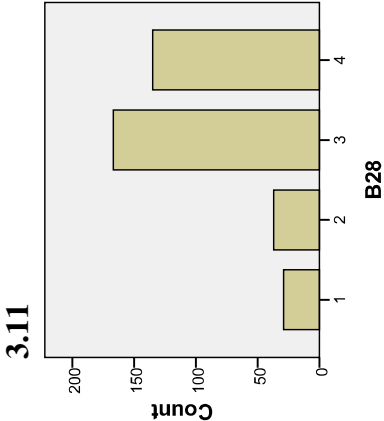
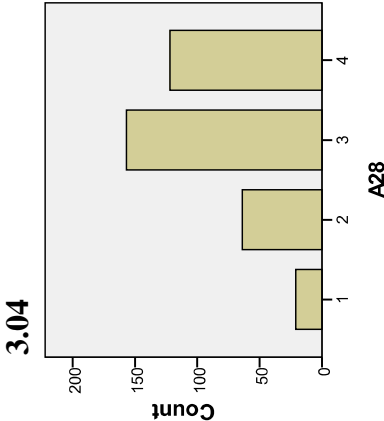




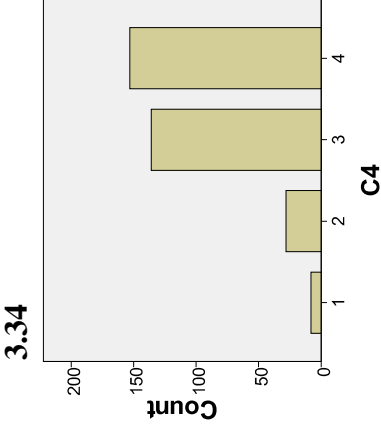
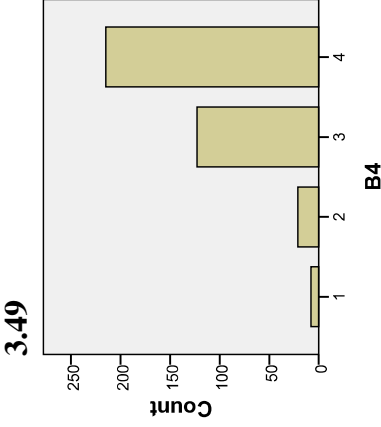
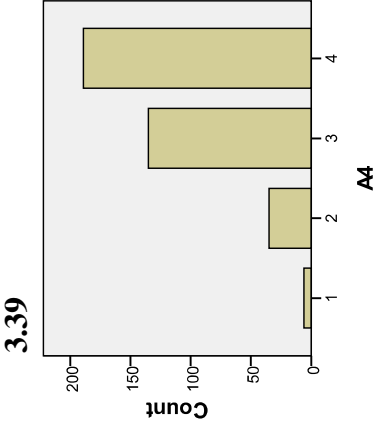
Question 23. I find it easy to make decisions.



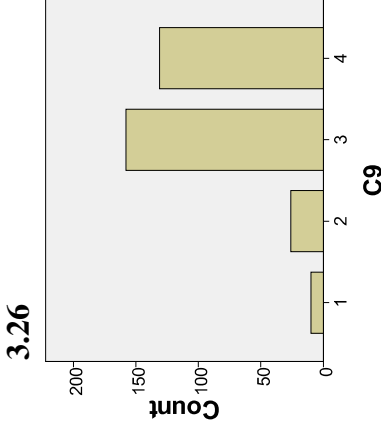
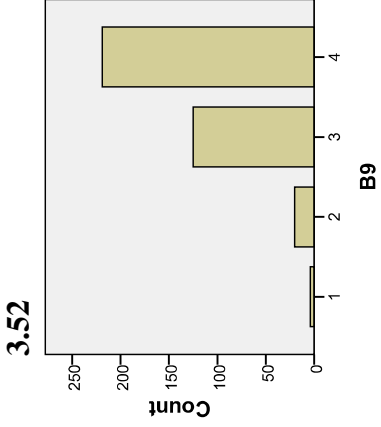
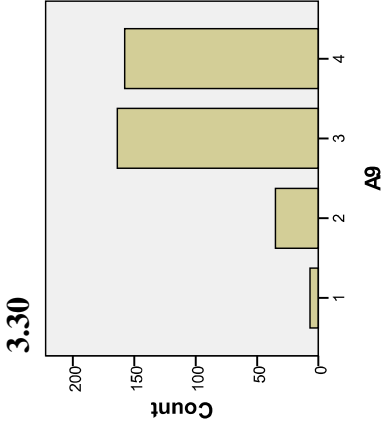
Question 28. I stay calm when things go wrong.



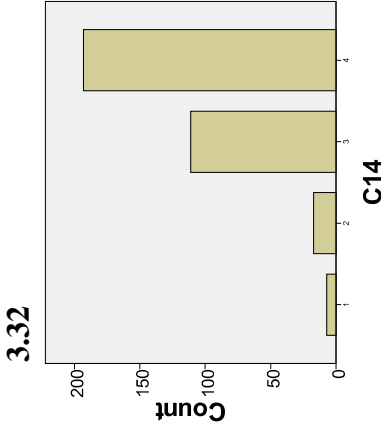
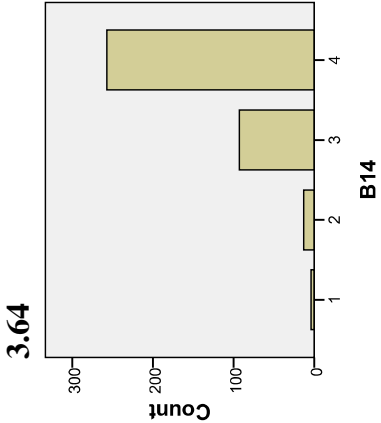
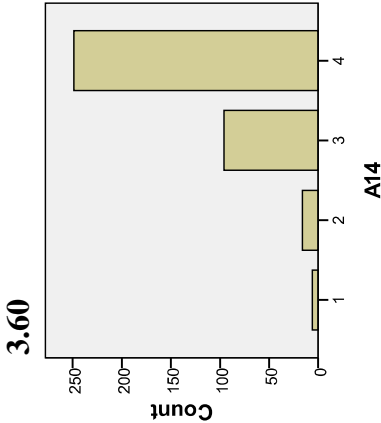
Question 4. I am good at making new friends.



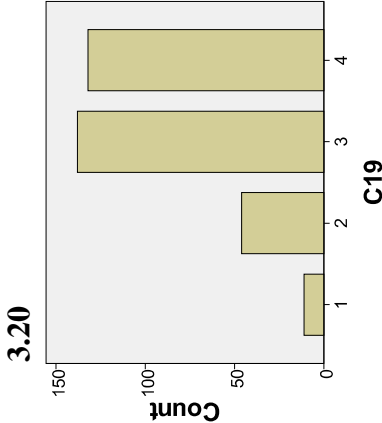
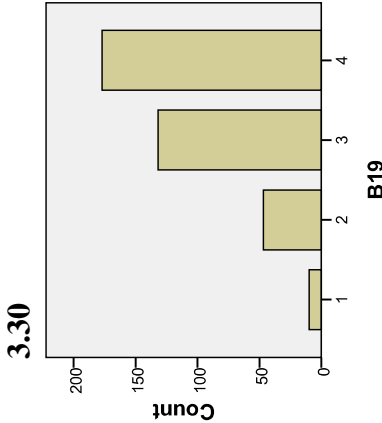
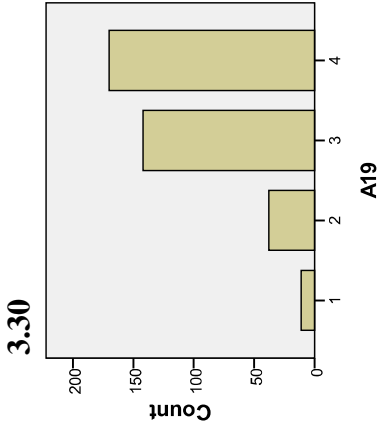
Question 9. I am happy to help other people get their jobs done.



Question 14. I like to listen to my friends' stories.

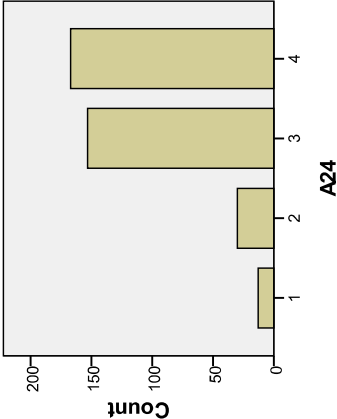


Question 19. I find it interesting that different people have different abilities.

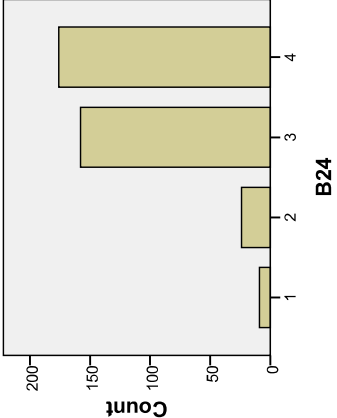


Question 24. I always listen carefully to instructions.

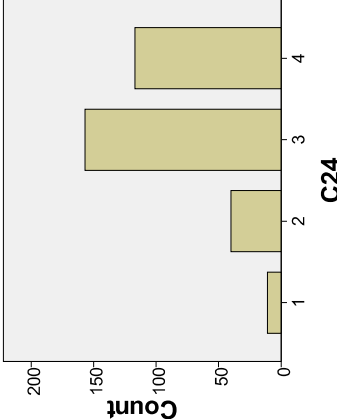
3.31



3.37

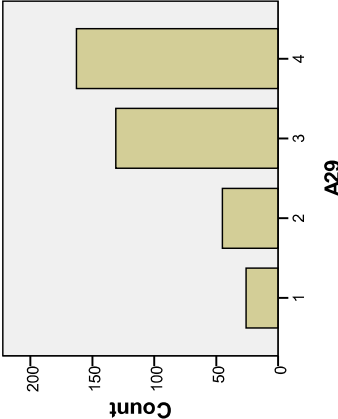


3.17

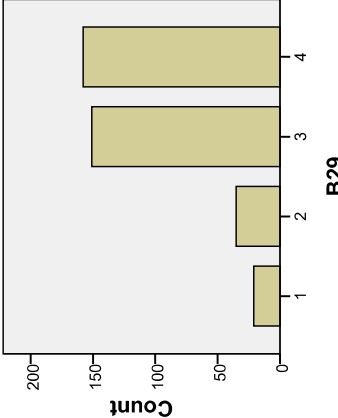


Question 29. I find it easy to talk with adults.

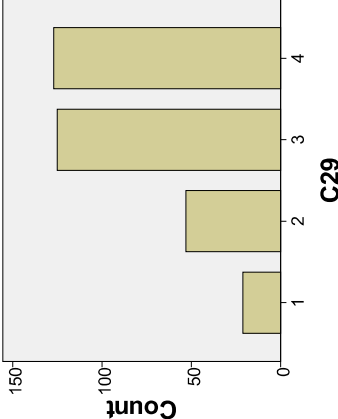
3.18



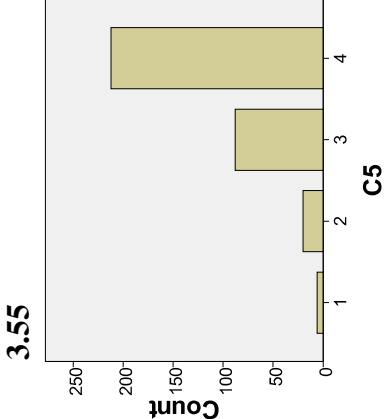
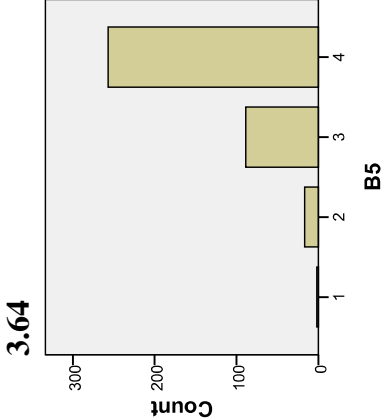
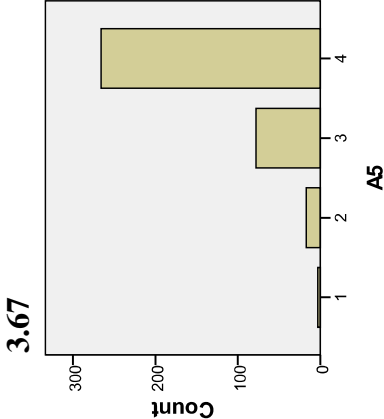
3.22



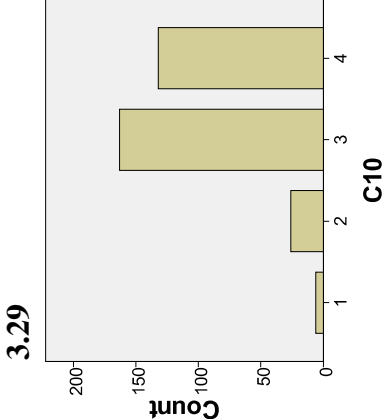
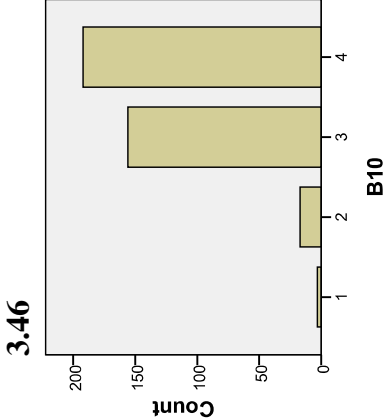
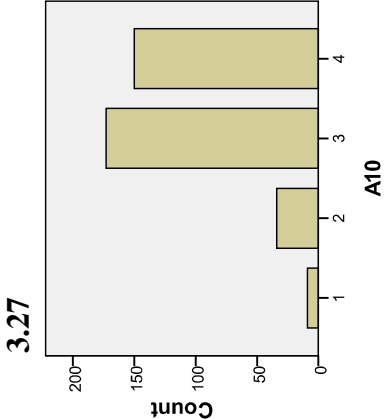
3.10



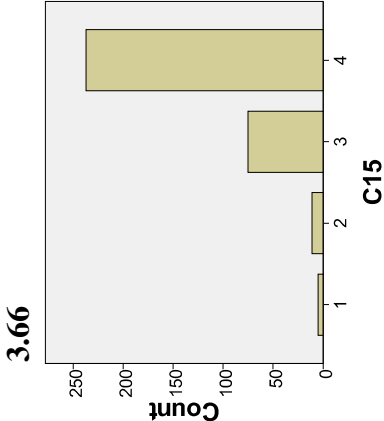
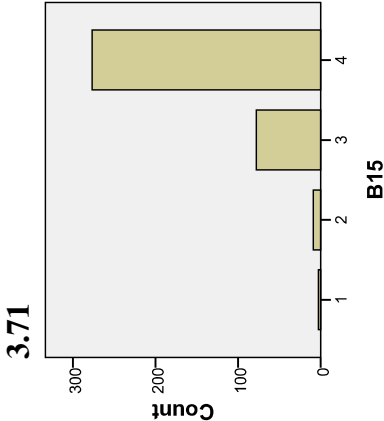
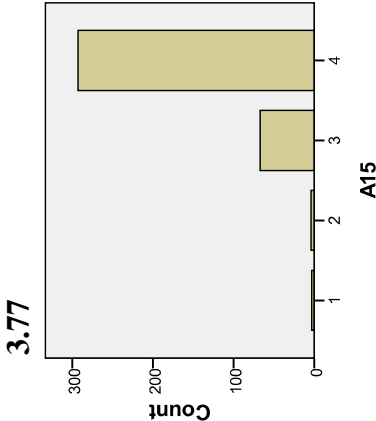
Question 5. I like team games.



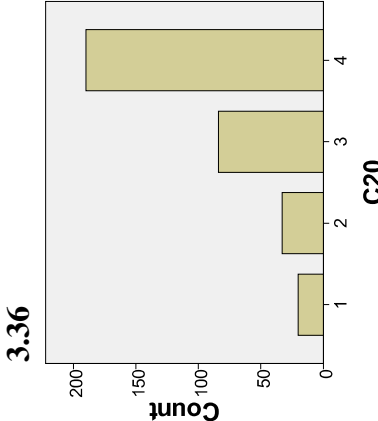
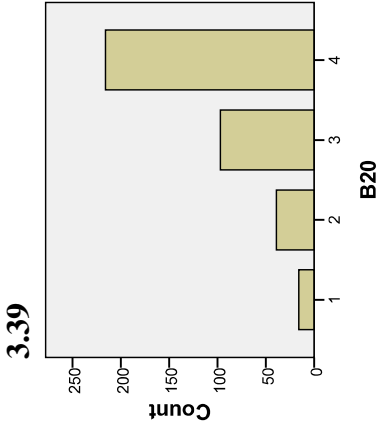
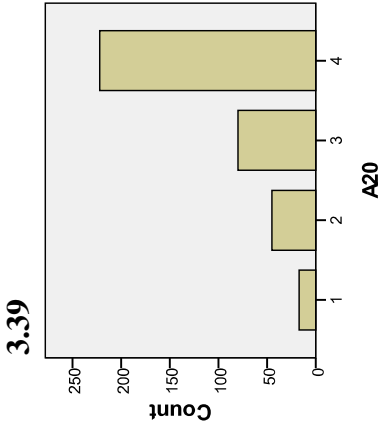
Question 10. I am good at listening to other peoples' points of view.



Question 15. I would support my team mates, even if we are losing.

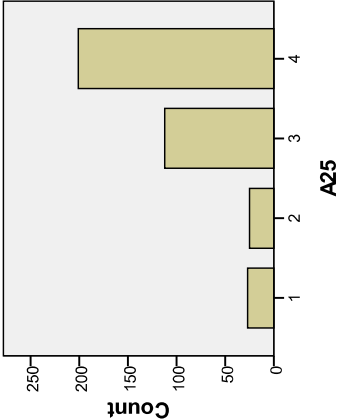


Question 20. I would be happy to be chosen as the leader of a team.

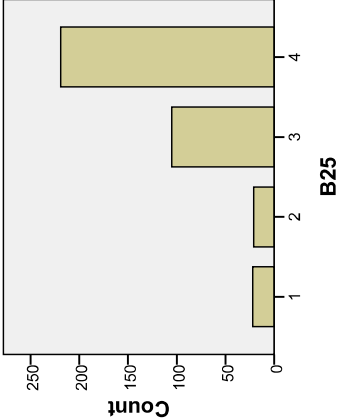


Question 25. I prefer working in a group to working on my own.

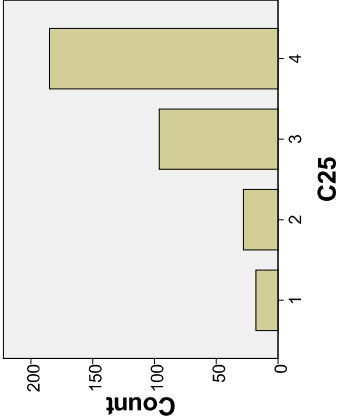
3.33



3.42

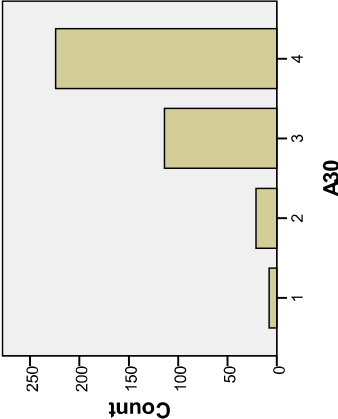


3.37

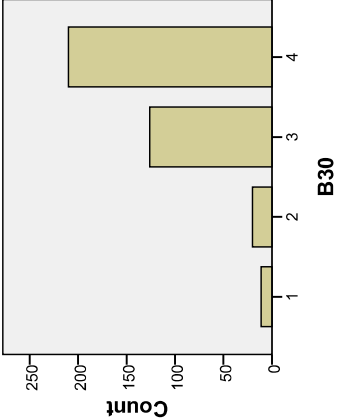


Question 30. I learn a lot from working in a group.

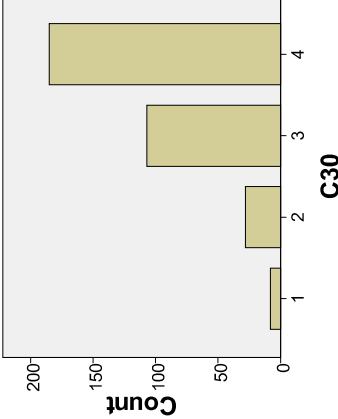
3.51



3.46

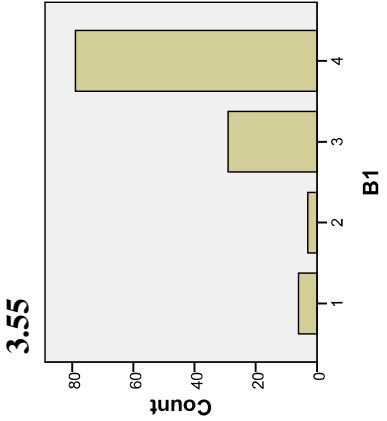
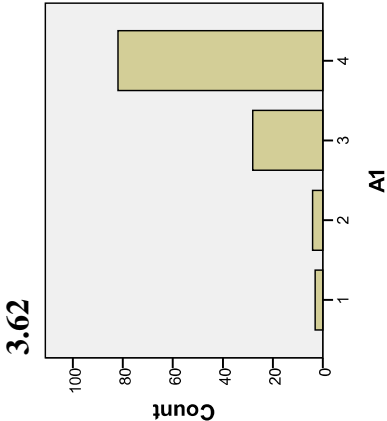


3.43

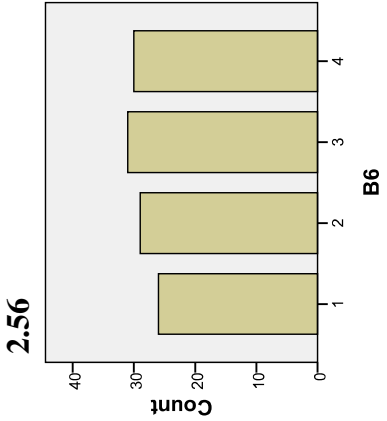
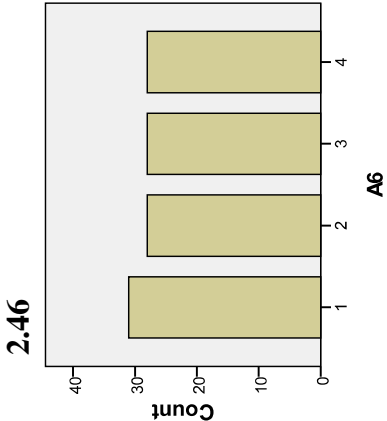


**CONTROL EXPERIMENT**

Question 1. I enjoy being outdoors.

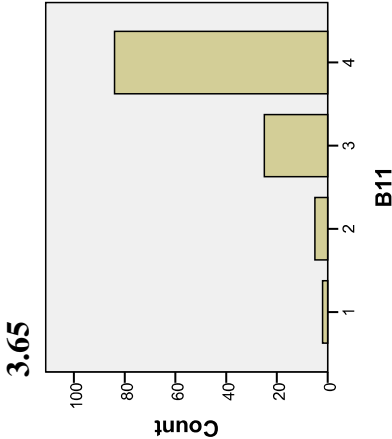
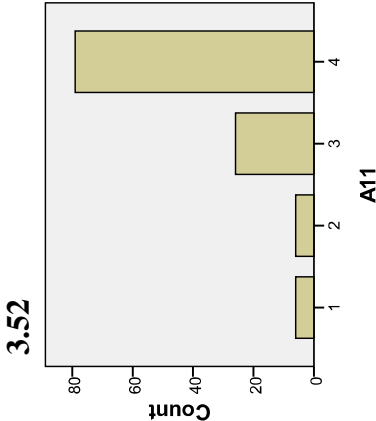


Question 6. I find nature programmes on the TV interesting.

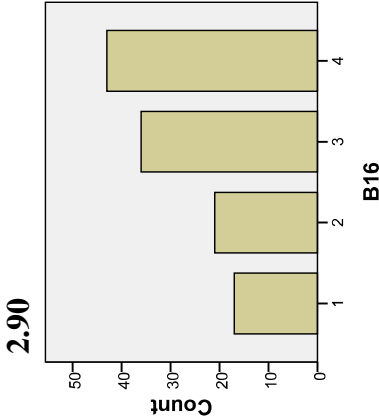
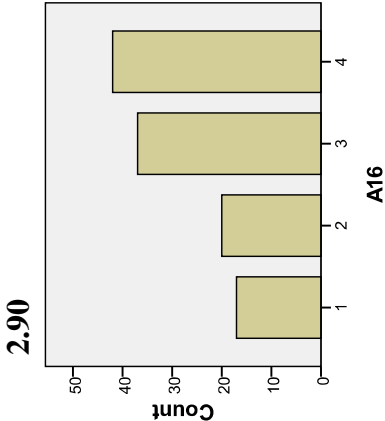




Question 11. I look forward to school classes that take us outdoors.

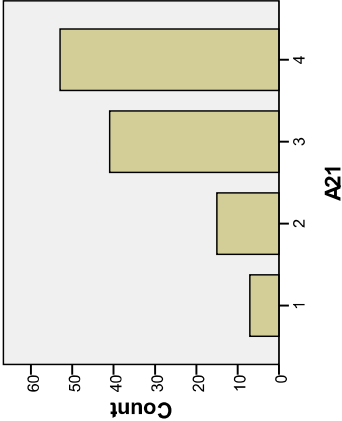


Question 16. I think it is interesting to explore the countryside.

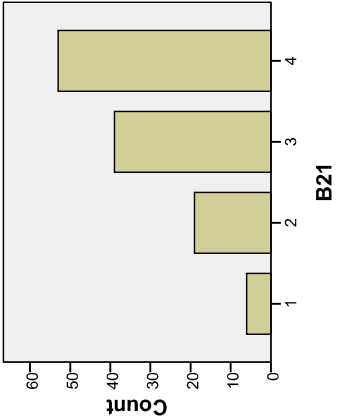


Question 21. I believe it is important to protect the outdoor environment.

3.21

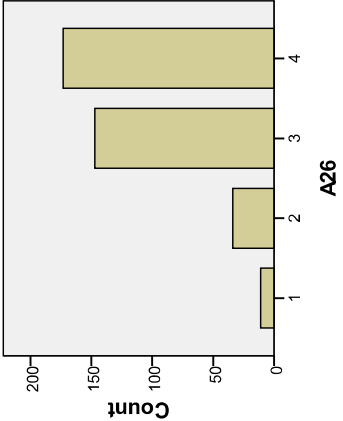


3.19

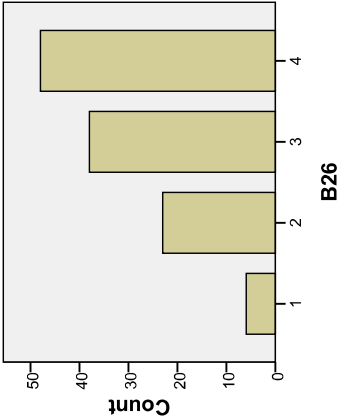


Question 26. I would like to protect the outdoor environment.

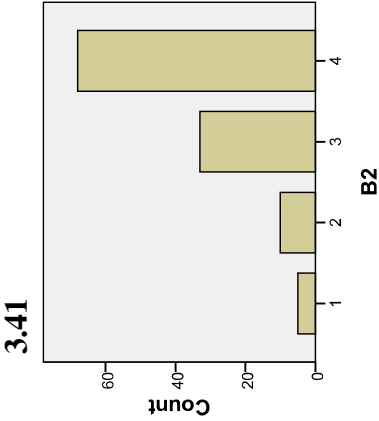
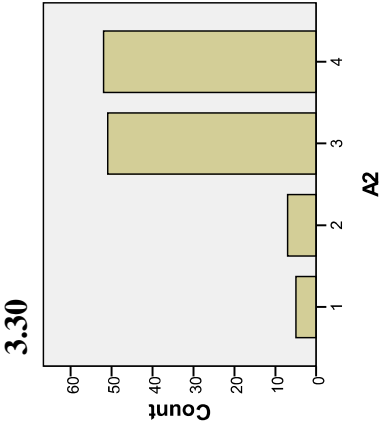
3.08



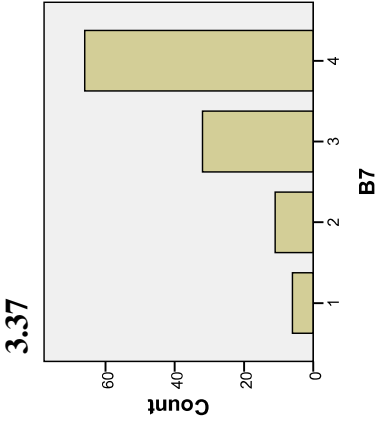
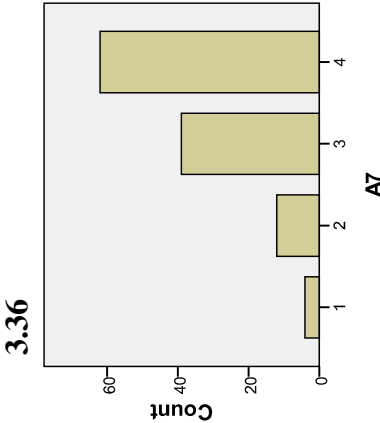
3.11



Question 2. I like to help with jobs around school.

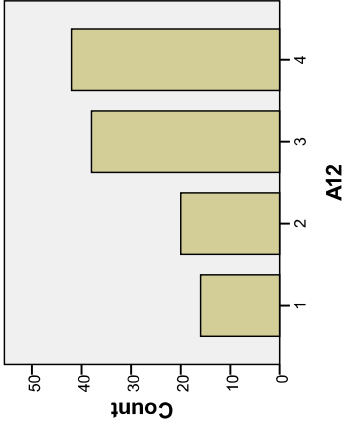


Question 7. I think it is important to have hobbies that keep you busy.

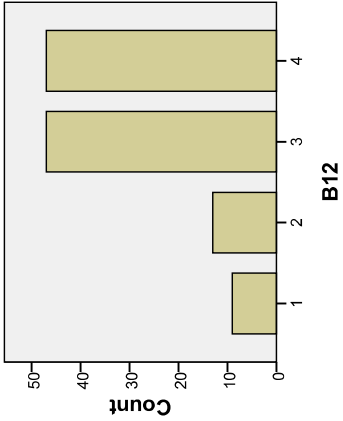


Question 12. I enjoy the challenge of a problem that is difficult.

2.91

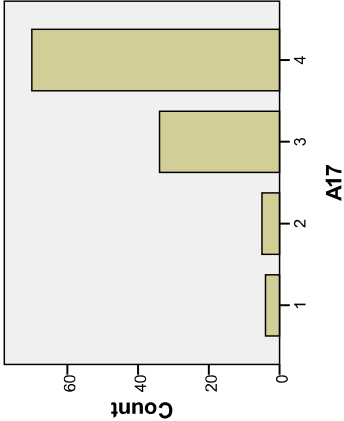


3.14

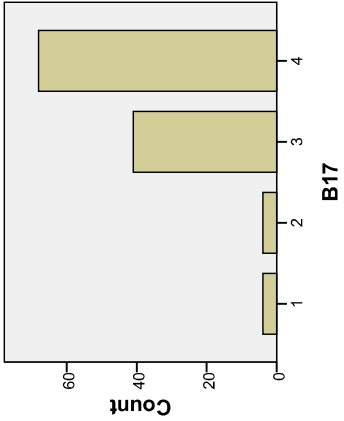


Question 17. I like to be active and get on with things.

3.50

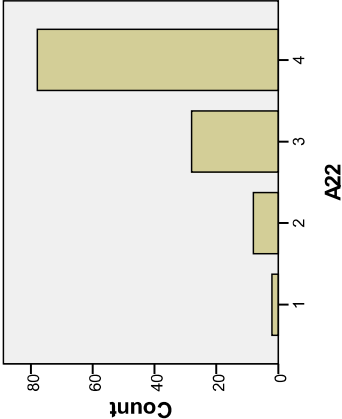


3.48

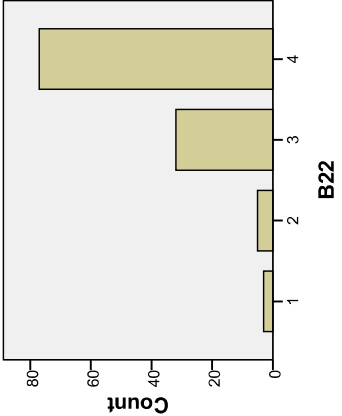


Question 22. I feel proud about the tasks I have successfully completed.

3.57

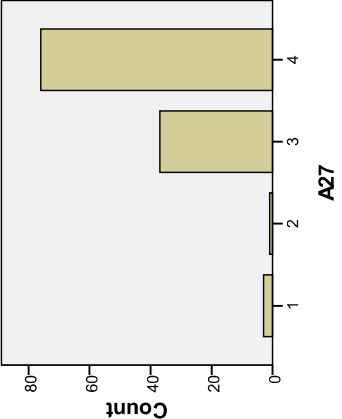


3.56

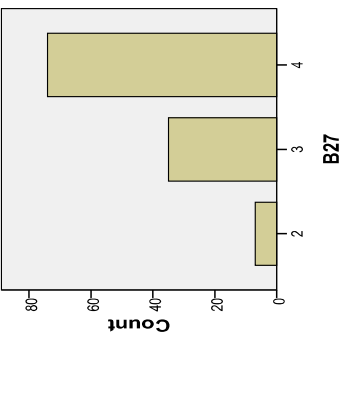


Question 27. I always try to do my best.

3.59

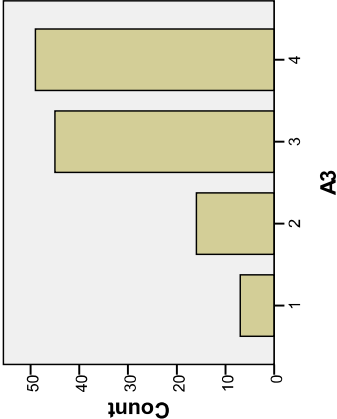


3.58



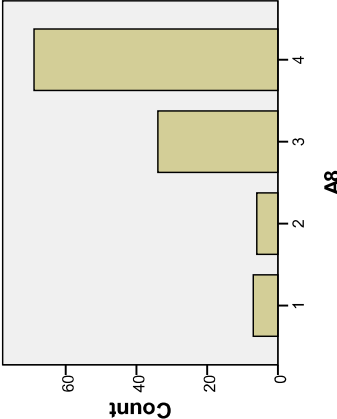
Question 3. I am a confident person.

3.16

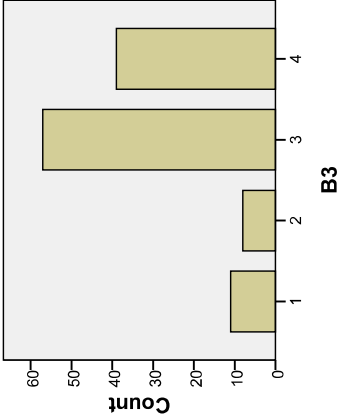


Question 8. I enjoy trying new experiences.

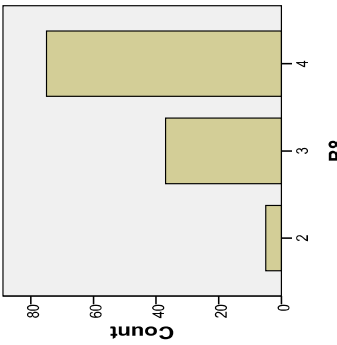
3.42



3.08

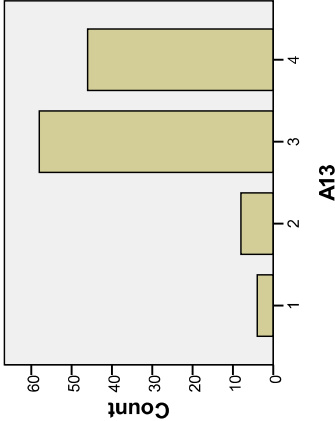


3.60

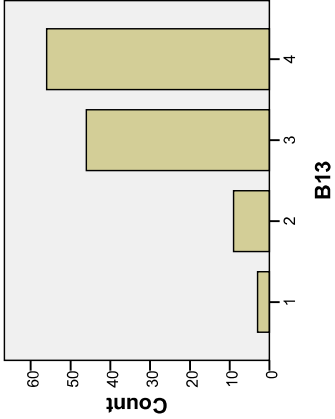


Question 13. I have the ability to finish what I set out to do.

3.26

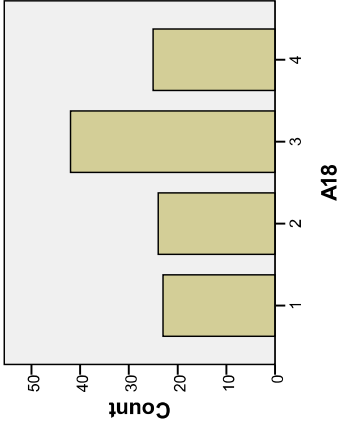


3.36

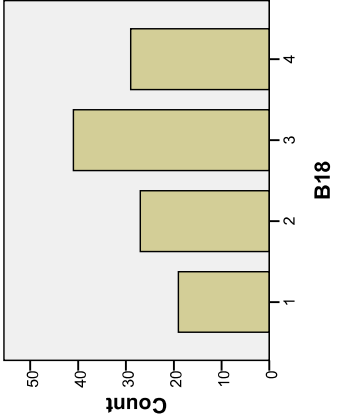


Question 18. I look forward to presenting my work to my class.

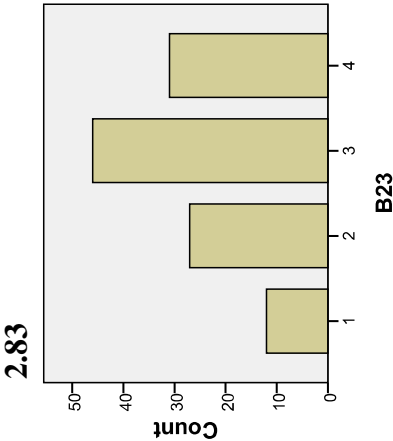
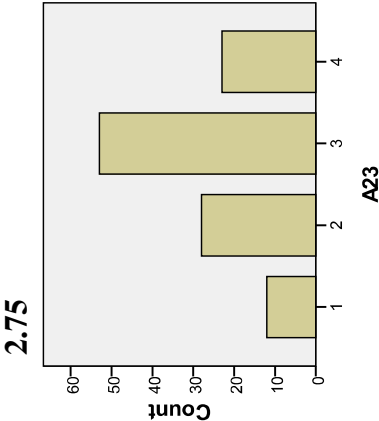
2.61



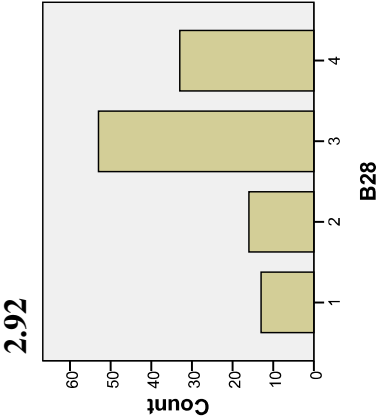
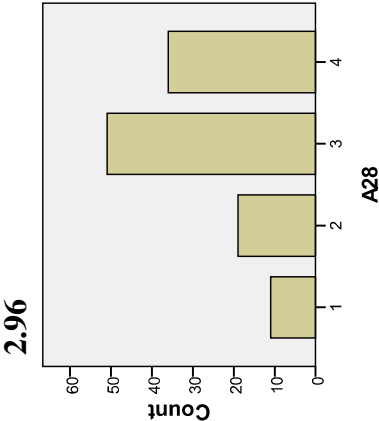
2.69



Question 23. I find it easy to make decisions.



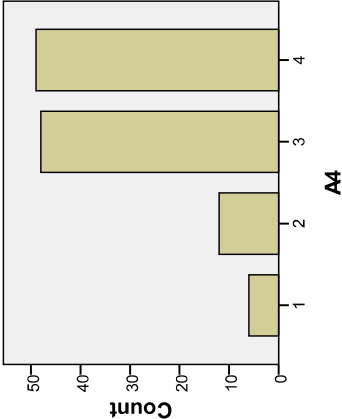
Question 28. I stay calm when things go wrong.



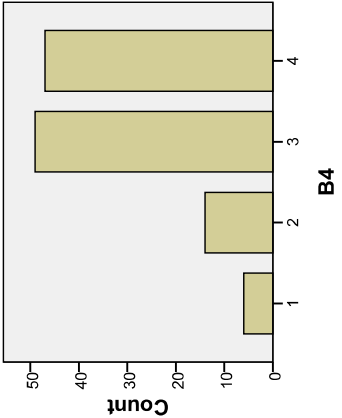


Question 4. I am good at making new friends.

3.22

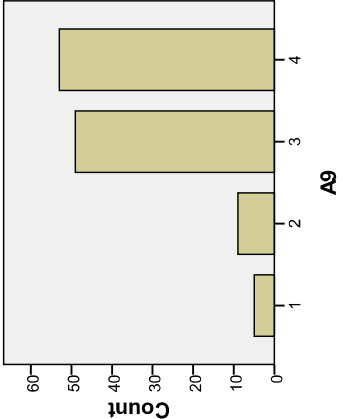


3.18

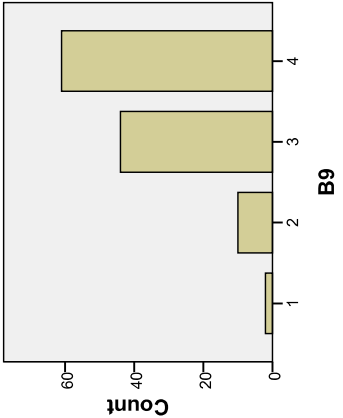


Question 9. I am happy to help other people get their jobs done.

3.29

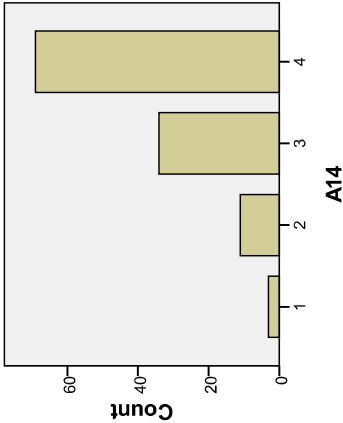


3.40

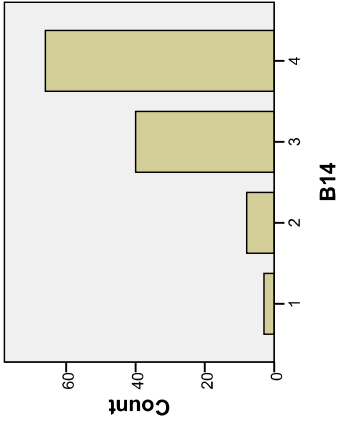


Question 14. I like to listen to my friends' stories.

3.44

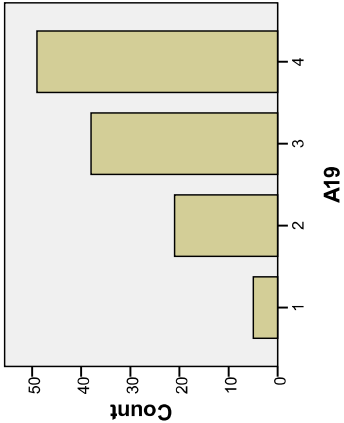


3.44

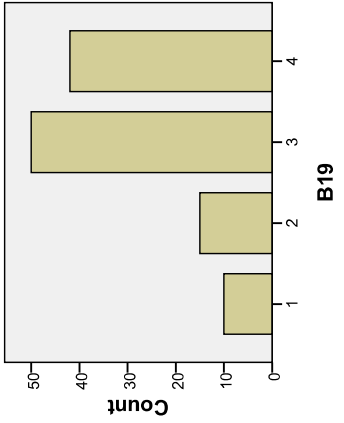


Question 19. I find it interesting that different people have different abilities.

3.16

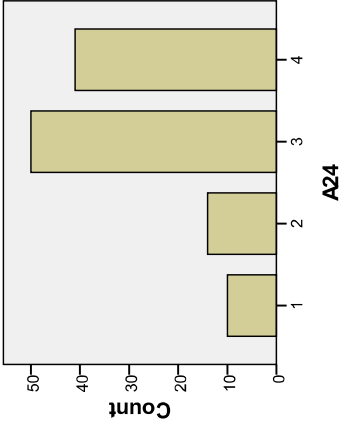


3.06

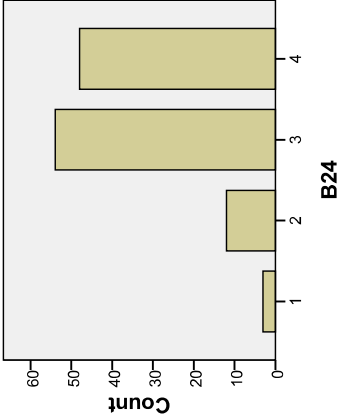


Question 24. I always listen carefully to instructions.

3.28

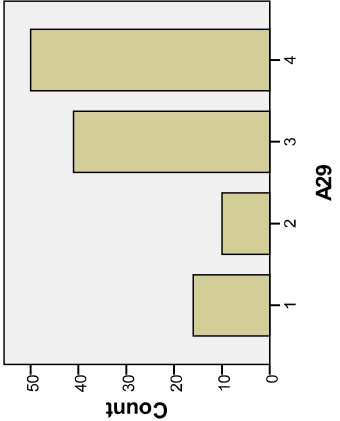


3.29

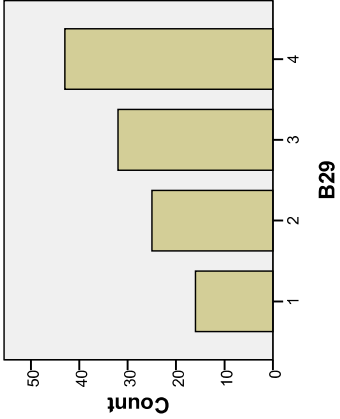


Question 29. I find it easy to talk with adults.

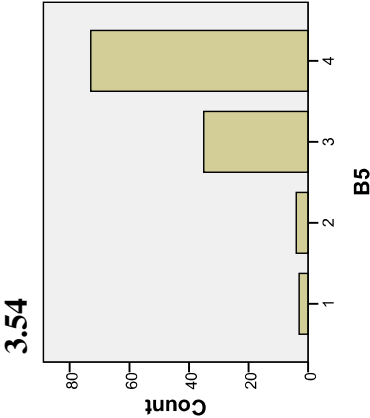
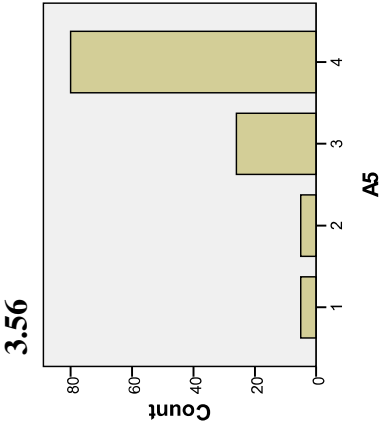
3.07



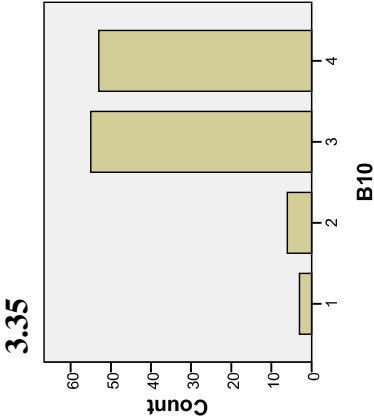
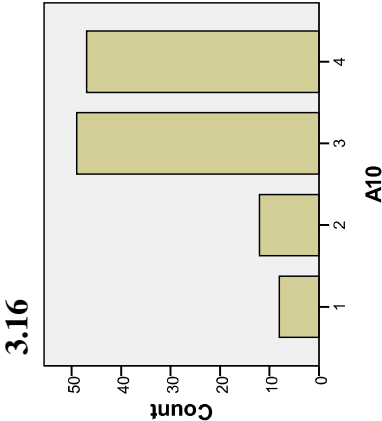
2.88



Question 5. I like team games.

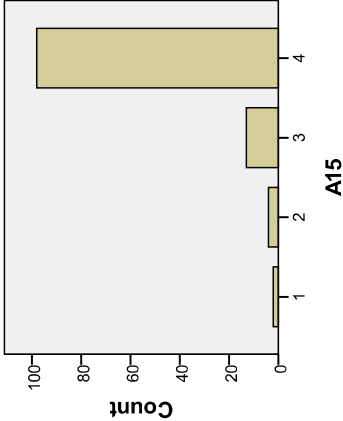


Question 10. I am good at listening to other peoples' points of view.

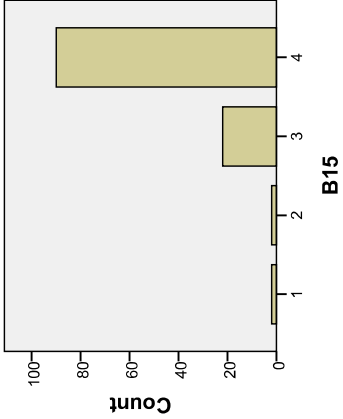


Question 15. I would support my team mates, even if we are losing.

3.77

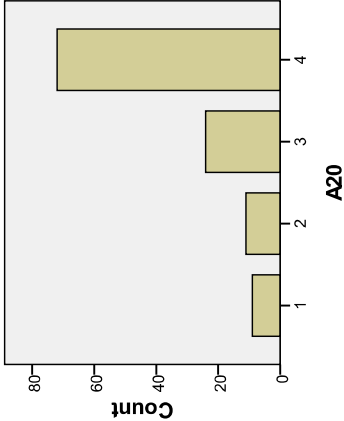


3.72

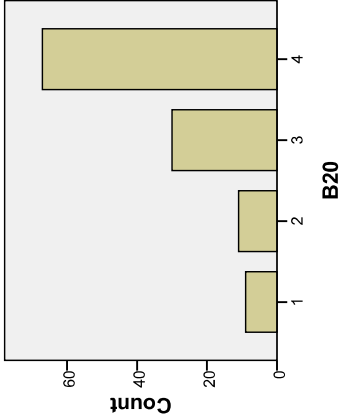


Question 20. I would be happy to be chosen as the leader of a team.

3.37

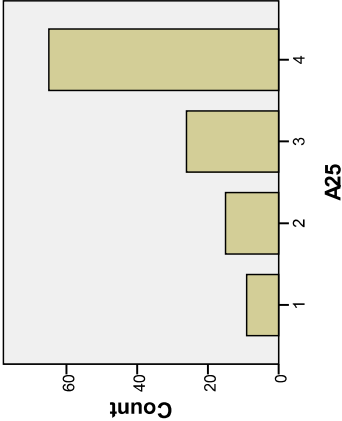


3.32

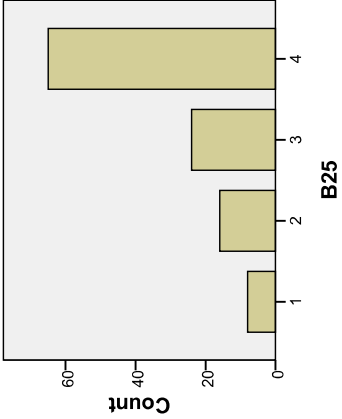


Question 25. I prefer working in a group to working on my own.

3.28

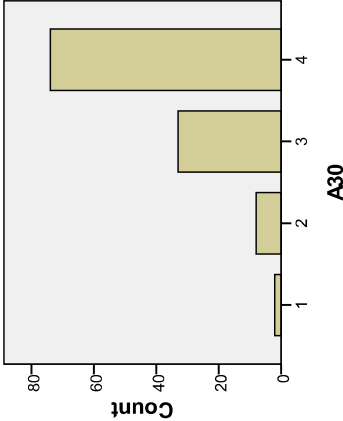


3.29

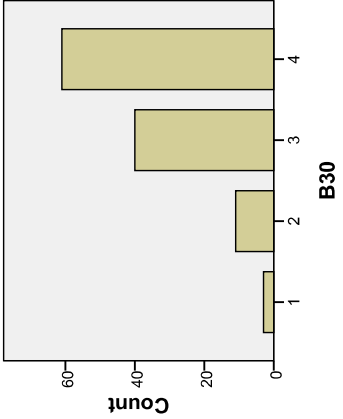


Question 30. I learn a lot from working in a group.

3.53



3.38



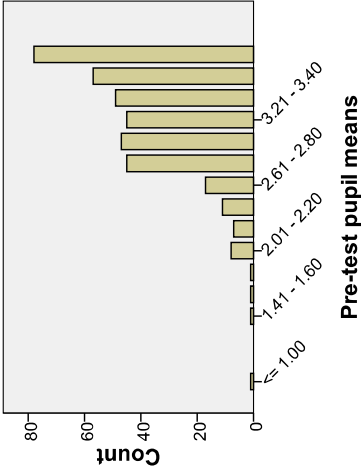
### **Appendix 3.2    Histograms for the distribution of pupils means scores in factors.**

Histograms are presented for the Main Study and the Control Experiment. In each case the data for each factor are for the whole sample and then for boys and girls separately. The mean score is bottom-left of the chart in each case.

For the Main Study there are histograms for the Pre-test (368 cases: 184 boys, 184 girls), Post-test (368 cases: 184 boys, 184 girls) and Delayed Test (328 cases: 169 boys, 159 girls). For the Control Experiment there are histograms for the Pre-test and the Post-test (117 cases: 68 boys, 49 girls).

N.B. both horizontal and vertical scales change with sample size and distribution of mean scores.

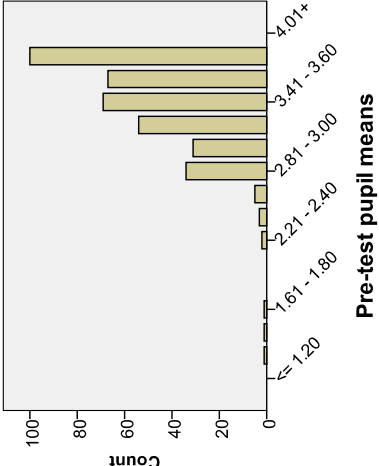
**MAIN STUDY – Whole sample**  
 Factor: Awareness of the outdoor environment



Mean = 3.33

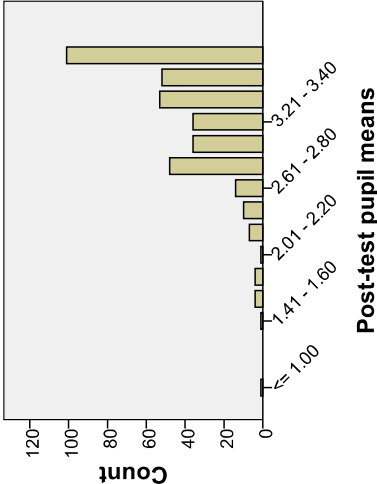
Pre-test pupil means

Factor: Perception of Competence



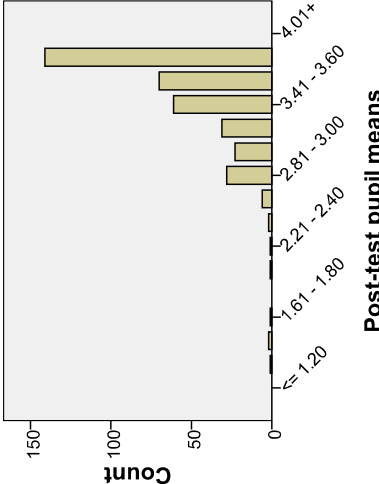
Mean = 3.50

Pre-test pupil means



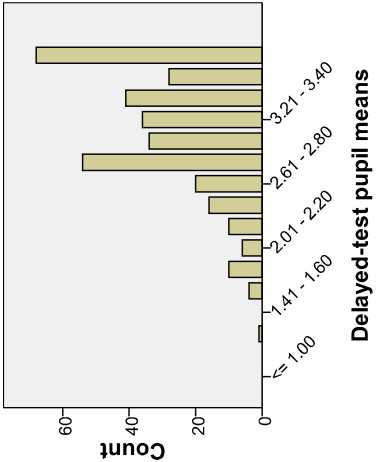
Mean = 3.38

Post-test pupil means



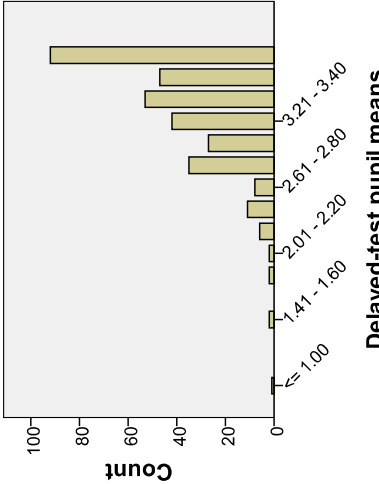
Mean = 3.57

Post-test pupil means



Mean = 3.22

Delayed-test pupil means

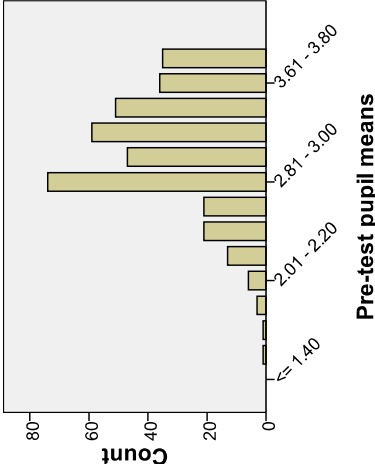


Mean = 3.41

Delayed-test pupil means



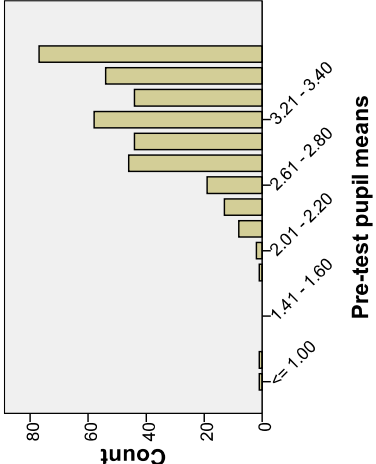
Factor: Perception of Confidence



Pre-test pupil means

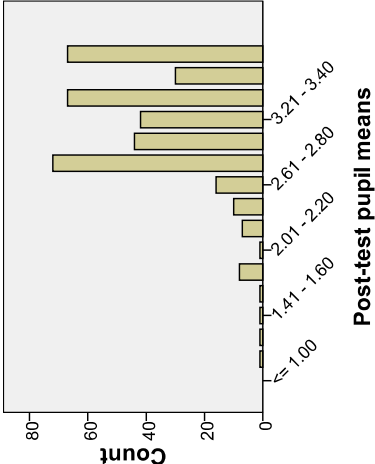
Mean = 3.18

Factor: Awareness of Others



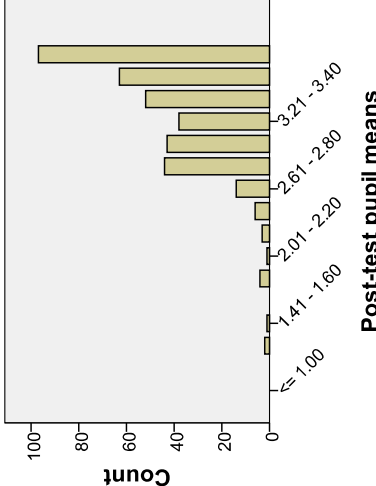
Pre-test pupil means

Mean = 3.35



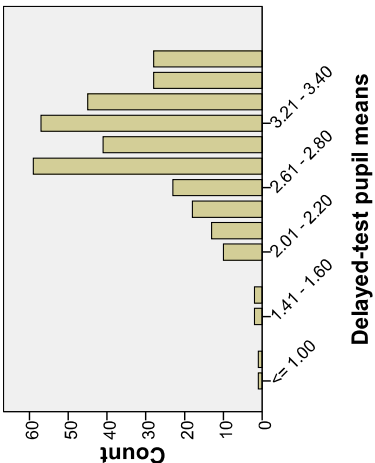
Post-test pupil means

Mean = 3.28



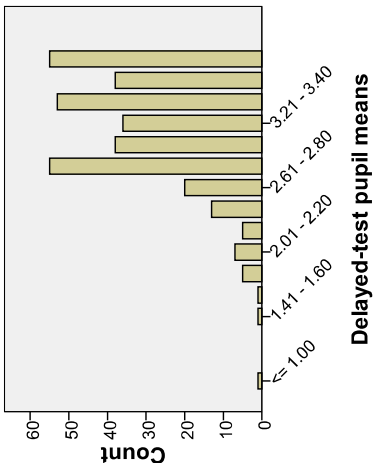
Post-test pupil means

Mean = 3.42



Delayed-test pupil means

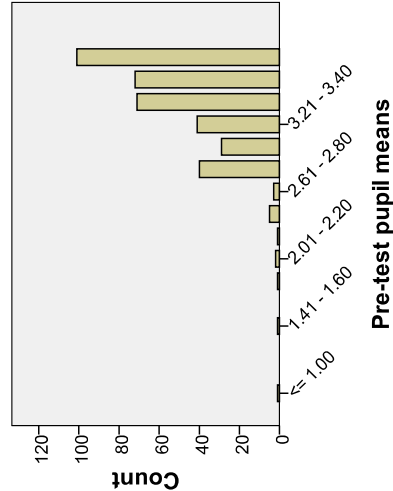
Mean = 3.14



Delayed-test pupil means

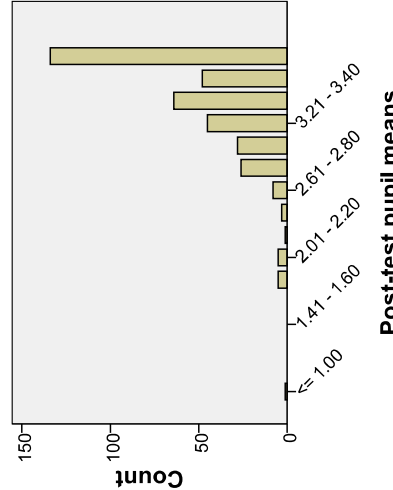
Mean = 3.26

Factor: Attitude towards Groups



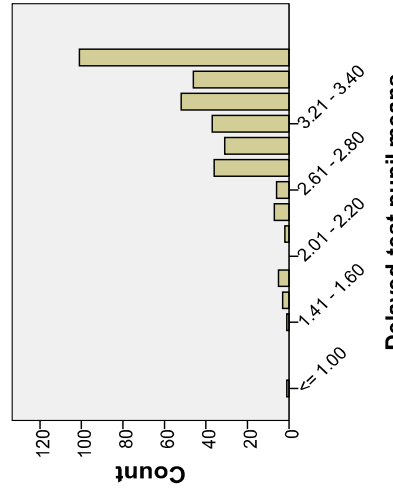
Pre-test pupil means

Mean = 3.49



Post-test pupil means

Mean = 3.51

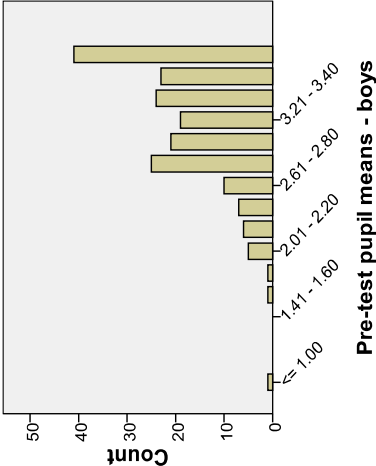


Delayed-test pupil means

Mean = 3.44

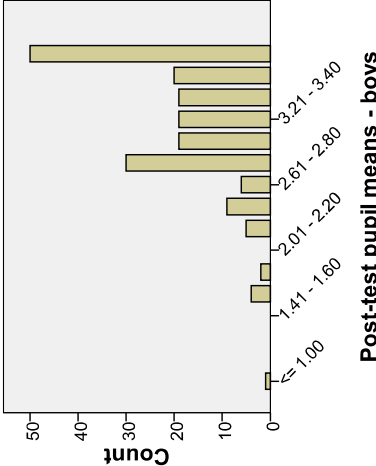
**MAIN STUDY – By gender**

Factor: Awareness of the outdoor environment



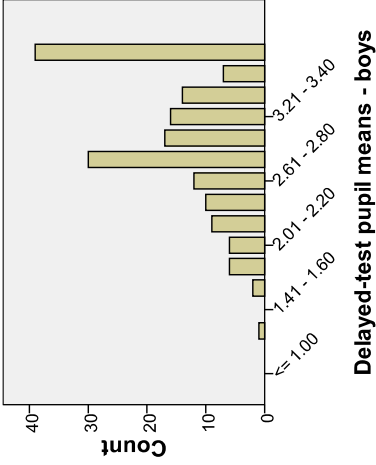
**Pre-test pupil means - boys**

Mean = 3.29



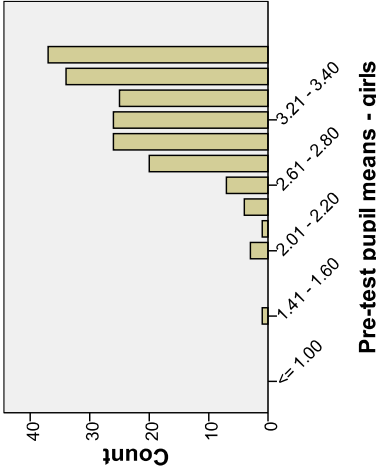
**Post-test pupil means - boys**

Mean = 3.31



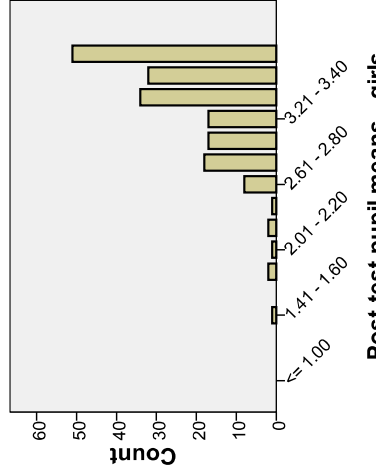
**Delayed-test pupil means - boys**

Mean = 3.14



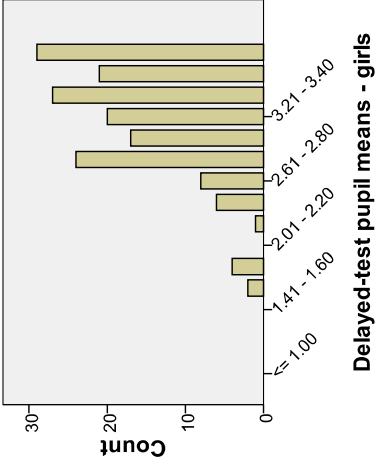
**Pre-test pupil means - girls**

Mean = 3.38



**Post-test pupil means - girls**

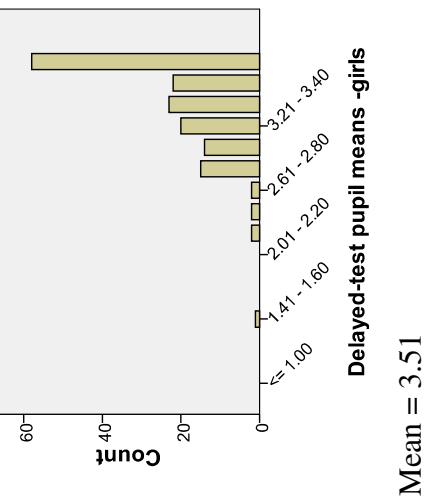
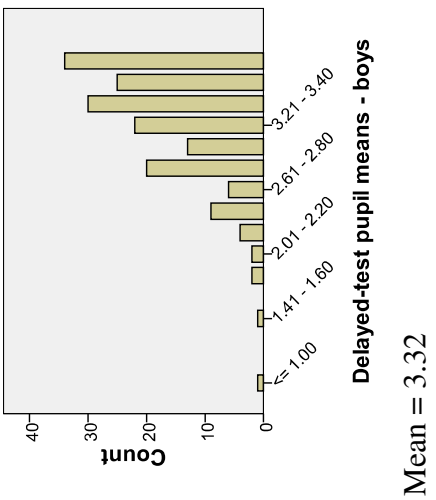
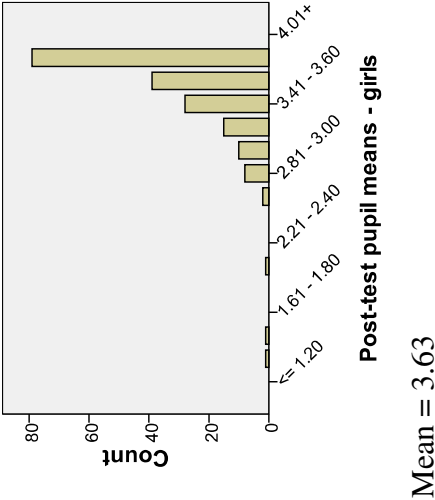
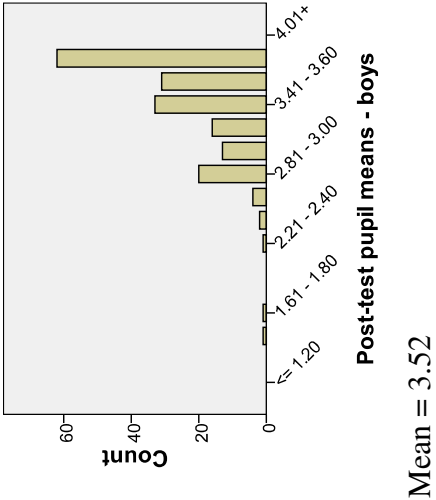
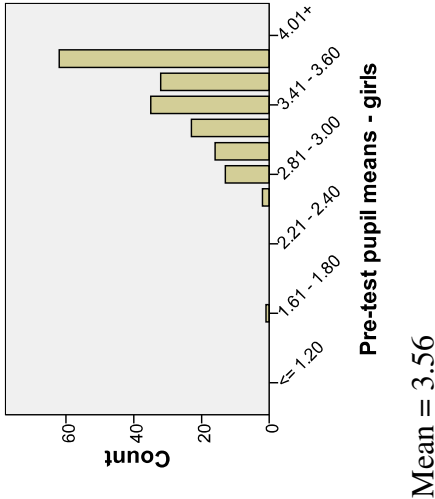
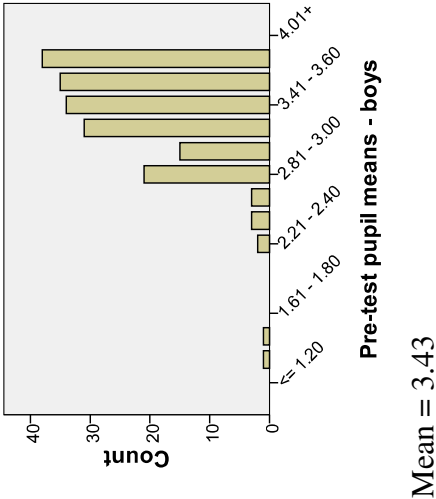
Mean = 3.45



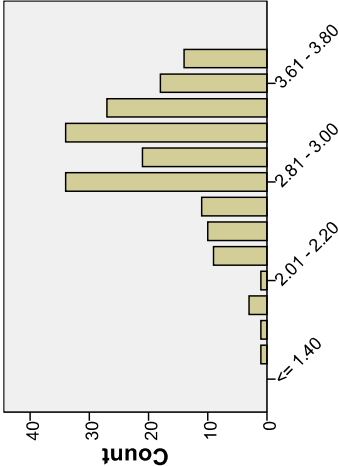
**Delayed-test pupil means - girls**

Mean = 3.31

Factor: Perception of Competence

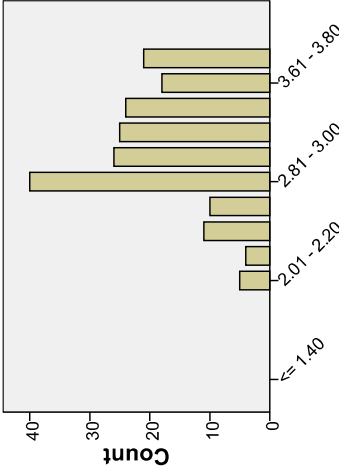


Factor: Perception of Confidence



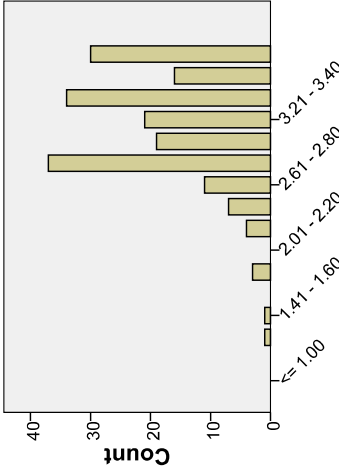
Pre-test pupil means - boys

Mean = 3.16



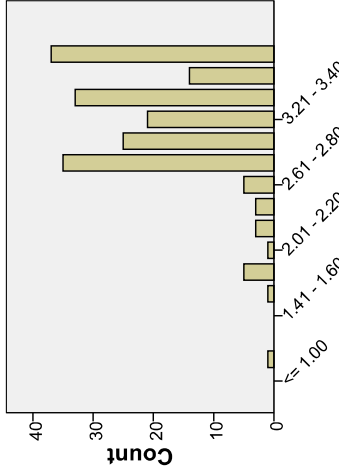
Pre-test pupil means - girls

Mean = 3.20



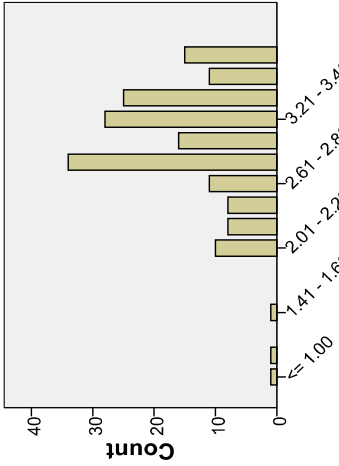
Post-test pupil means - boys

Mean = 3.26



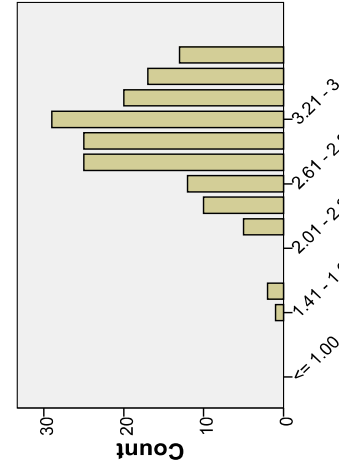
Post-test pupil means - girls

Mean = 3.30



Delayed-test pupil means - boys

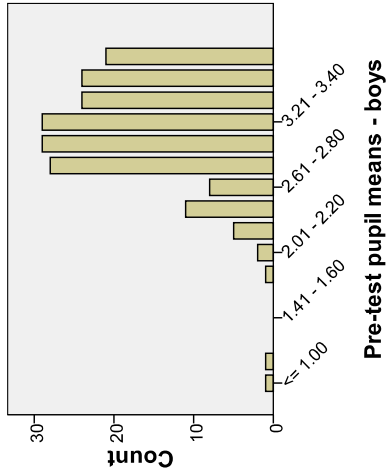
Mean = 3.10



Delayed-test pupil means - girls

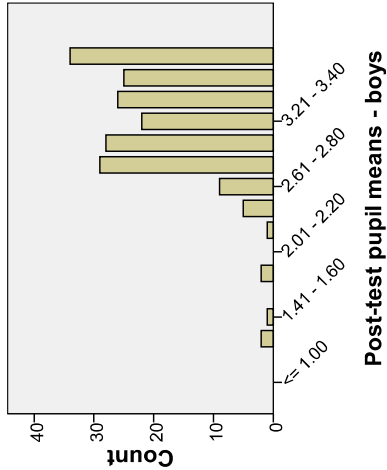
Mean = 3.18

Factor: Awareness of Others



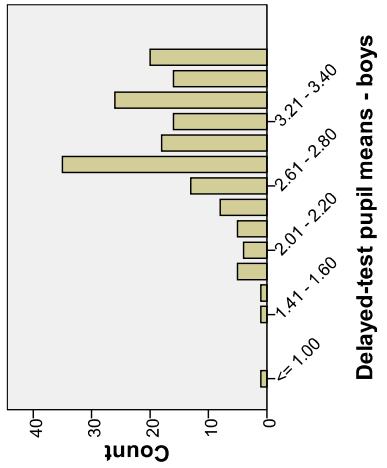
Pre-test pupil means - boys

Mean = 3.23



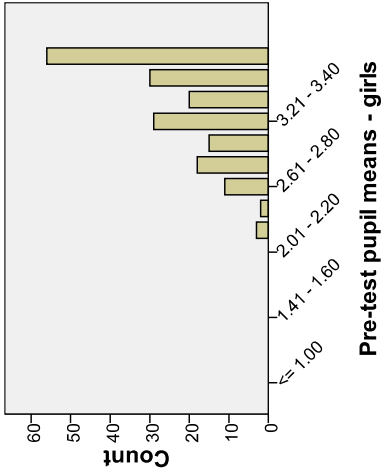
Post-test pupil means - boys

Mean = 3.31



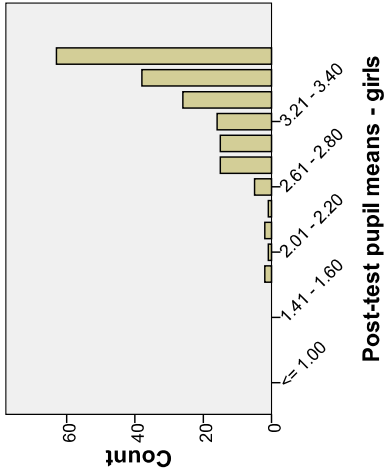
Delayed-test pupil means - boys

Mean = 3.13



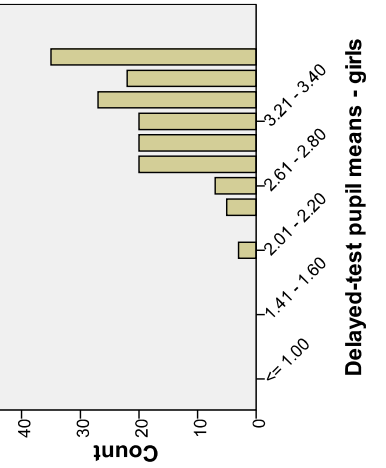
Pre-test pupil means - girls

Mean = 3.46



Post-test pupil means - girls

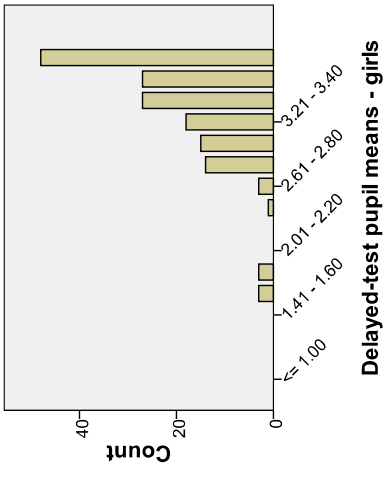
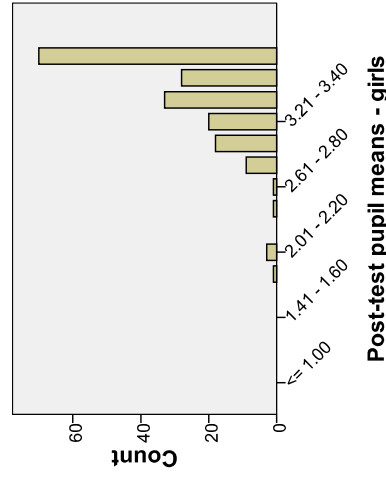
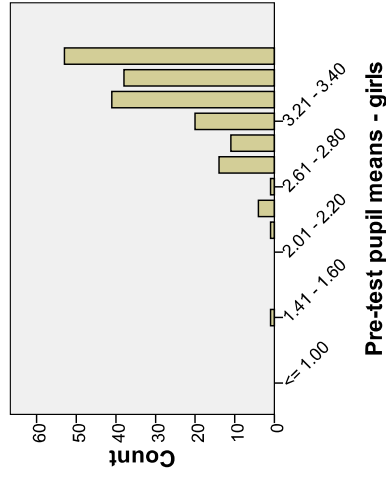
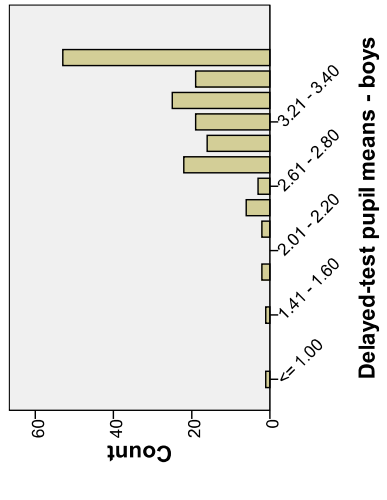
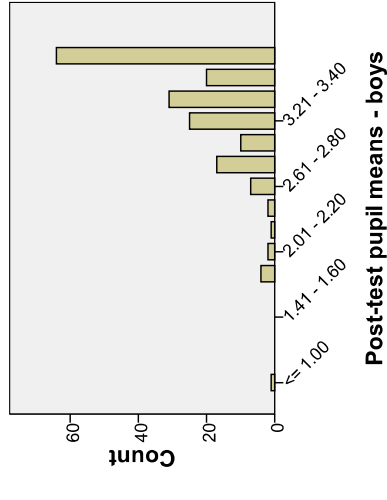
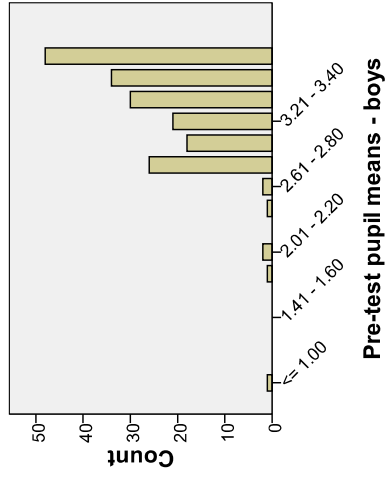
Mean = 3.53



Delayed-test pupil means - girls

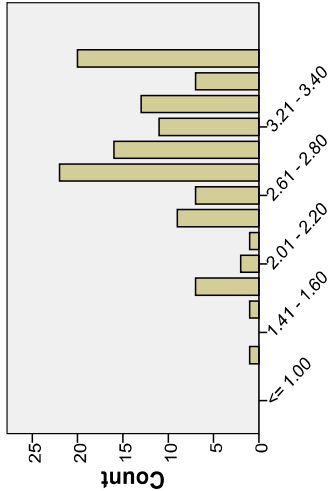
Mean = 3.40

Factor: Attitude towards Groups



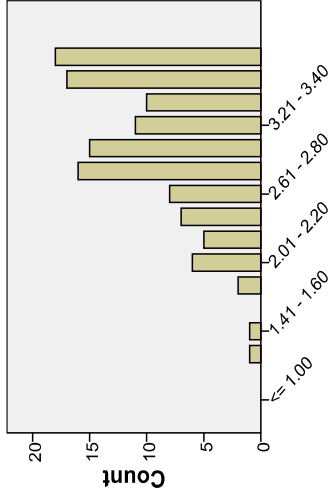
# CONTROL EXPERIMENT – Whole sample

Factor: Awareness of the outdoor environment



Pre-test pupil means

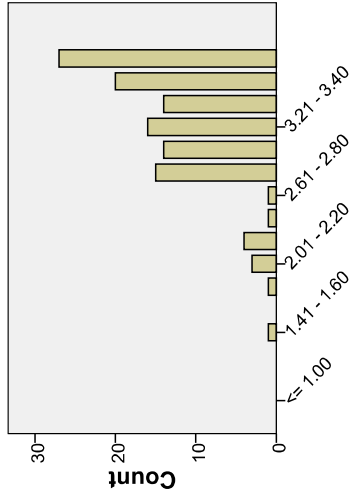
Mean = 3.13



Post-test pupil means

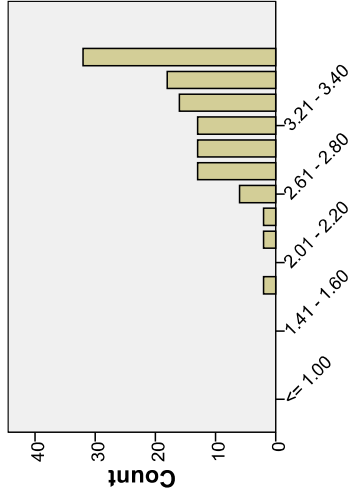
Mean = 3.16

Factor: Perception of Competence



Pre-test pupil means

Mean = 3.37

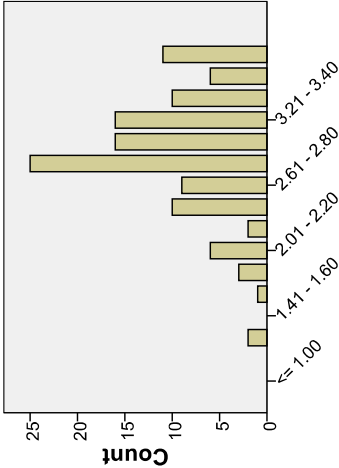


Post-test pupil means

Mean = 3.43

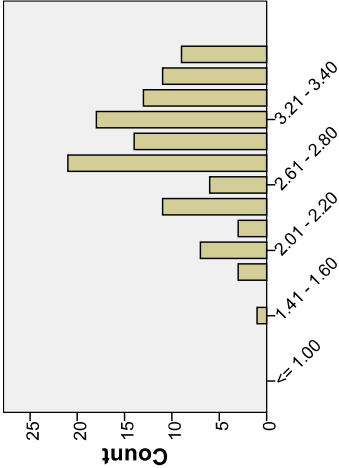


Factor: Perception of Confidence



Pre-test pupil means

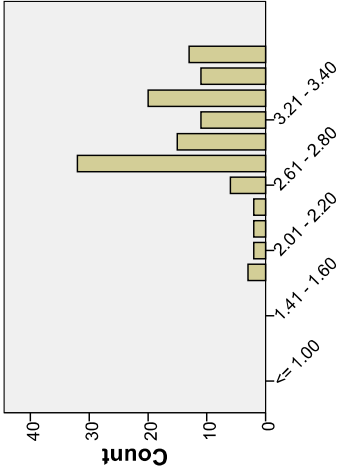
Mean = 3.03



Post-test pupil means

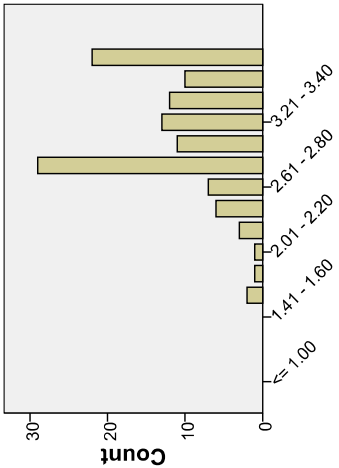
Mean = 3.08

Factor: Awareness of Others



Pre-test pupil means

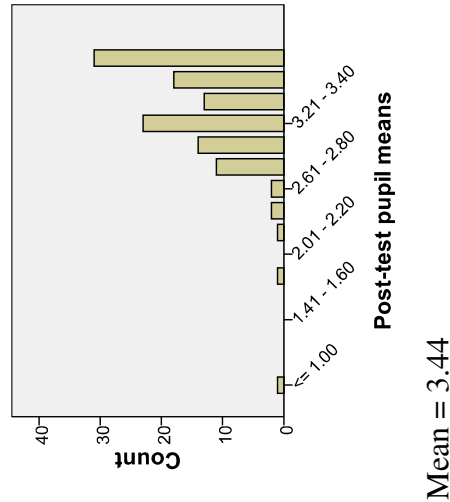
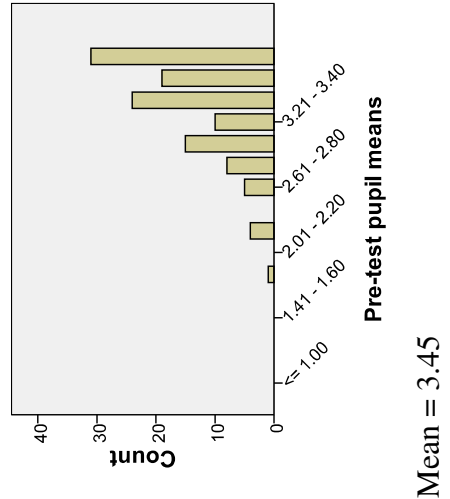
Mean = 3.21



Post-test pupil means

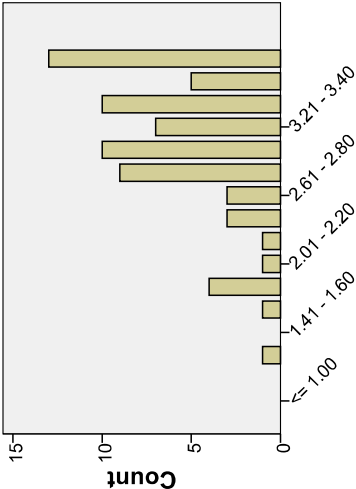
Mean = 3.21

Factor: Attitude towards Groups



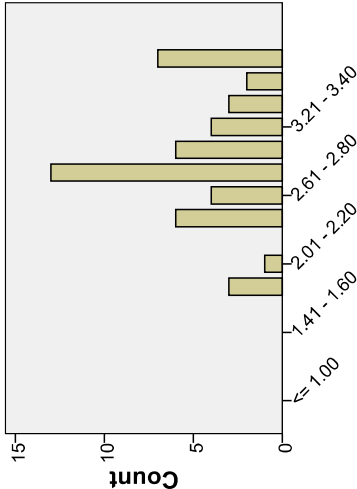
**CONTROL EXPERIMENT – By gender**

Factor: Awareness of the Outdoor Environment



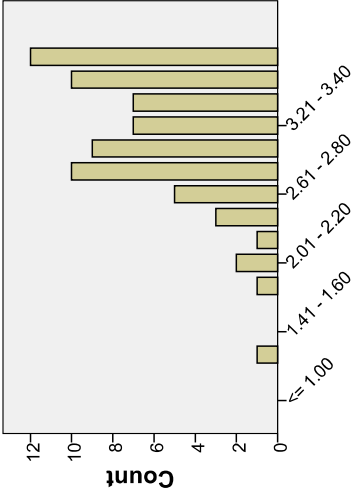
Pre-test pupil means - boys

Mean = 3.18



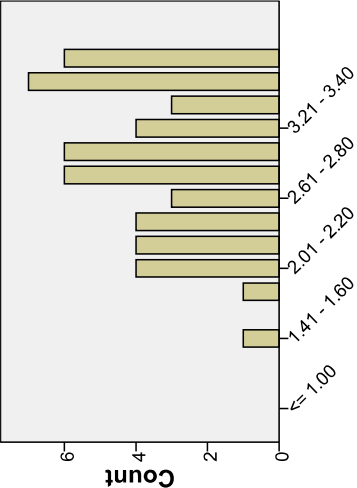
Pre-test pupil means - girls

Mean =3.06



Post-test pupil means - boys

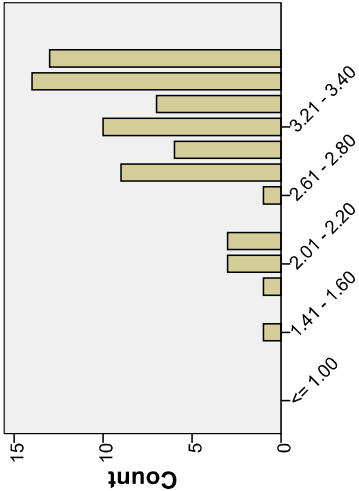
Mean = 3.24



Post-test pupil means - girls

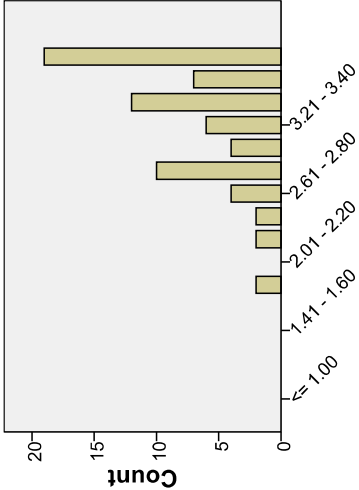
Mean = 3.04

Factor: Perception of Competence



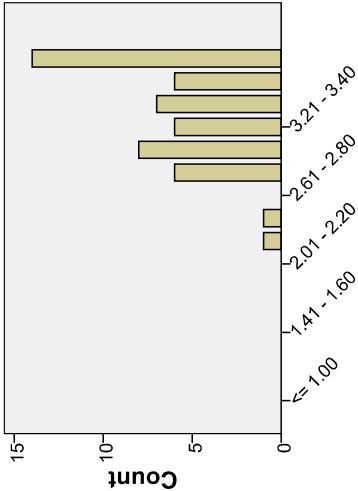
Pre-test pupil means - boys

Mean = 3.31



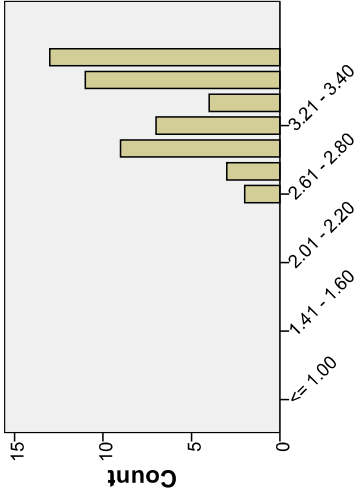
Post-test pupil means - boys

Mean = 3.36



Pre-test pupil means - girls

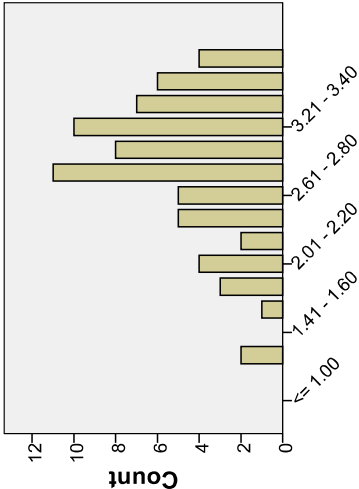
Mean = 3.46



Post-test pupil means - girls

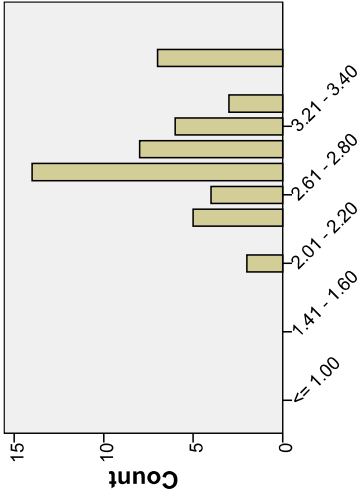
Mean = 3.52

Factor: Perception of Confidence



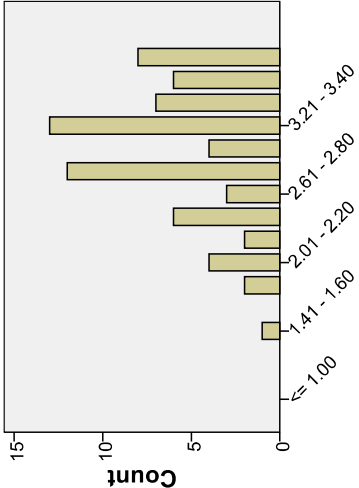
Post-test pupil means - boys

Mean = 2.98



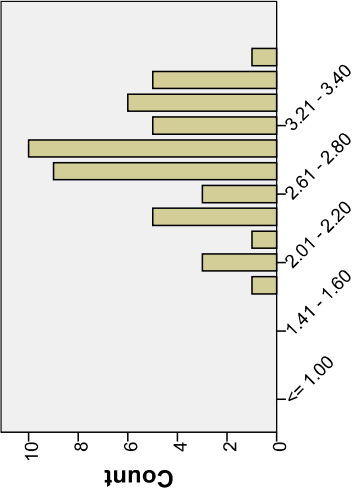
Post-test pupil means - girls

Mean = 3.10



Pre-test pupil means - boys

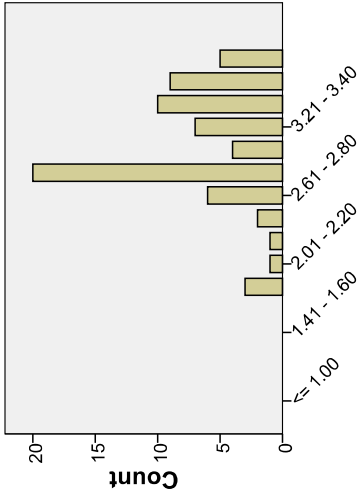
Mean = 3.10



Pre-test pupil means - girls

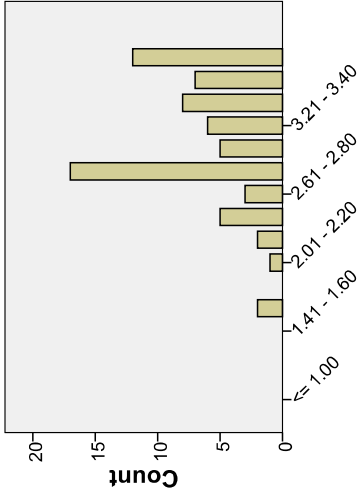
Mean = 3.04

Factor: Awareness of Others



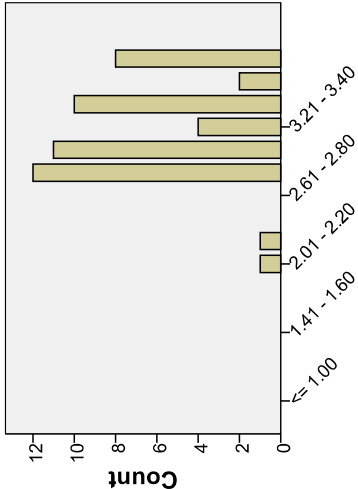
Pre-test pupil means - boys

Mean = 3.15



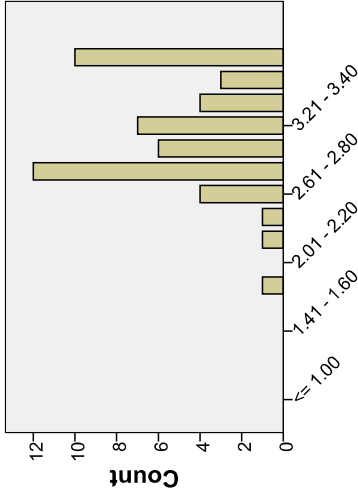
Post-test pupil means - boys

Mean = 3.18



Pre-test pupil means - girls

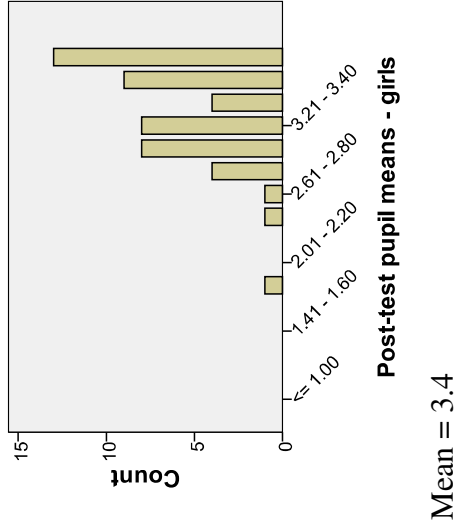
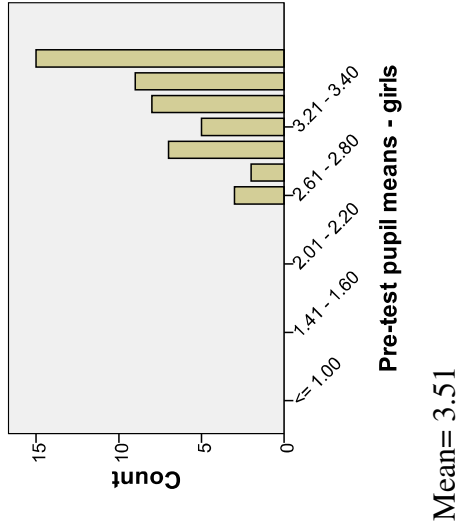
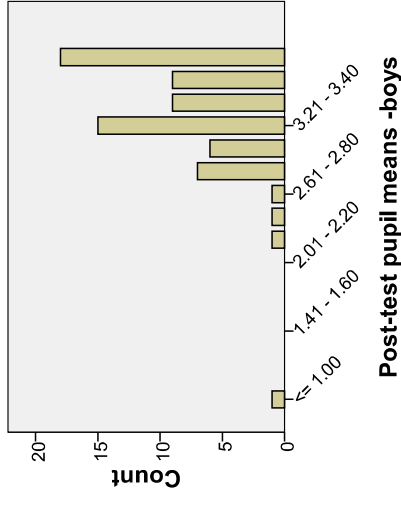
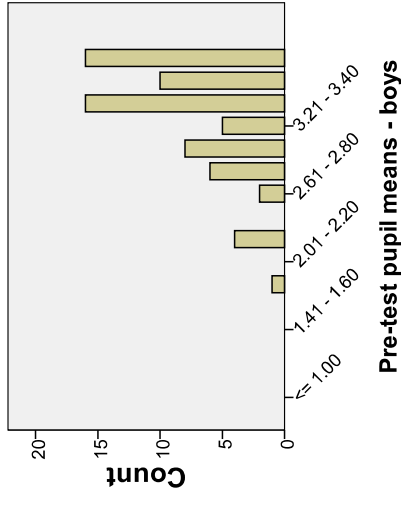
Mean = 3.29



Post-test pupil means - girls

Mean = 3.24

Factor: Attitude towards Groups



### Appendix 3.3 Effect sizes and statistical significance of differences in pupils' mean scores in factors

#### Notes

1. For the mean scores on which these tests are based, see Appendix 3.2 Histograms for the distribution of pupils' mean scores in factors.

2.  $r$  is Effect size, calculated using:

$$r = z / \sqrt{N}, \quad \text{where } N \text{ is the total number of observations.}$$

Statistically significant small (0.2) to medium (0.5) effect sizes are shown in bold type.

3.  $p$  is the two-tailed statistical significance: the significance level is taken to be 0.05.

4. Sample sizes are shown in brackets in the Sample column for "Awareness of Outdoor Environment" only, but apply to all factors.

#### Main Study

Factor: Awareness of Outdoor Environment

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases (368)	0.10/0.009	
	Boys (184)	0.05/0.309	
	Girls (184)	<b>0.15/0.005</b>	Post- > Pre-
<b>Difference in mean scores Pre-test – Delayed-test</b>	All (308)	<b>0.16/0.001</b>	Pre- > Delayed
	Boys (160)	<b>0.19/0.001</b>	Pre- > Delayed
	Girls (148)	0.13/0.031	
<b>Difference in mean scores Post-test – Delayed-test</b>	All (308)	<b>0.29/0.001</b>	Post- > Delayed
	Boys (160)	<b>0.30/0.001</b>	Post- > Delayed
	Girls (148)	<b>0.28/0.001</b>	Post- > Delayed
<b>Difference between genders, Pre-test</b>	Boys v Girls (184 v 184)	0.07/0.172	
<b>Difference between genders, Post-test</b>	Boys v Girls (184 v 184)	0.11/0.040	
<b>Difference between genders, Delayed-test</b>	Boys v Girls (169 v 159)	0.13/0.015	



Factor: Perception of Competence

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	<b>0.17</b> / <b>&lt;0.001</b>	Post- > Pre-
	Boys	<b>0.17</b> /0.001	Post- > Pre-
	Girls	<b>0.17</b> /0.001	Post- > Pre-
<b>Difference in mean scores Pre-test – Delayed-test</b>	All	<b>0.17</b> / <b>&lt;0.001</b>	Pre- > Delayed
	Boys	<b>0.23</b> / <b>&lt;0.001</b>	Pre- > Delayed
	Girls	0.10/0.089	
<b>Difference in mean scores Post-test – Delayed-test</b>	All	<b>0.30</b> / <b>&lt;0.001</b>	Post- > Delayed
	Boys	<b>0.34</b> / <b>&lt;0.001</b>	Post- > Delayed
	Girls	<b>0.26</b> / <b>&lt;0.001</b>	Post- > Delayed
<b>Difference between genders, Pre-test</b>	Boys v Girls	<b>0.16</b> /0.002	Girls > Boys
<b>Difference between genders, Post-test</b>	Boys v Girls	<b>0.15</b> /0.004	Girls > Boys
<b>Difference between genders, Delayed-test</b>	Boys v Girls	<b>0.19</b> /0.001	Girls > Boys

Factor: Perception of Confidence

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	<b>0.18</b> / <b>&lt;0.001</b>	Post- > Pre-
	Boys	<b>0.16</b> /0.002	Post- > Pre-
	Girls	<b>0.19</b> / <b>&lt;0.001</b>	Post- > Pre-
<b>Difference in mean scores Pre-test – Delayed-test</b>	All	0.08/0.039	
	Boys	0.14/0.012	
	Girls	0.02/0.770	
<b>Difference in mean scores Post-test – Delayed-test</b>	All	<b>0.26</b> / <b>&lt;0.001</b>	Post- > Delayed
	Boys	<b>0.31</b> / <b>&lt;0.001</b>	Post- > Delayed
	Girls	<b>0.20</b> /0.001	Post- > Delayed
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.02/0.702	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.05/0.370	
<b>Difference between genders, Delayed-test</b>	Boys v Girls	0.07/0.222	

Factor: Awareness of Others

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	<b>0.15</b> / <b>&lt;0.001</b>	Post- > Pre-
	Boys	0.14/0.009	
	Girls	<b>0.16</b> /0.002	Post- > Pre-
<b>Difference in mean scores Pre-test – Delayed-test</b>	All	<b>0.16</b> / <b>&lt;0.001</b>	Pre- > Delayed
	Boys	<b>0.18</b> /0.001	Pre- > Delayed
	Girls	0.14/0.017	
<b>Difference in mean scores Post-test – Delayed-test</b>	All	<b>0.28</b> / <b>&lt;0.001</b>	Post- > Delayed
	Boys	<b>0.31</b> / <b>&lt;0.001</b>	Post- > Delayed
	Girls	<b>0.26</b> / <b>&lt;0.001</b>	Post- > Delayed
<b>Difference between genders, Pre-test</b>	Boys v Girls	<b>0.25</b> / <b>&lt;0.001</b>	Girls > Boys
<b>Difference between genders, Post-test</b>	Boys v Girls	<b>0.25</b> / <b>&lt;0.001</b>	Girls > Boys
<b>Difference between genders, Delayed-test</b>	Boys v Girls	<b>0.24</b> / <b>&lt;0.000</b>	Girls > Boys

Factor: Attitude towards Groups

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	0.05/0.156	
	Boys	0.04/0.494	
	Girls	0.07/0.182	
<b>Difference in mean scores Pre-test – Delayed-test</b>	All	0.10/0.015	
	Boys	0.10/0.075	
	Girls	0.10/0.094	
<b>Difference in mean scores Post-test – Delayed-test</b>	All	0.14/0.001	
	Boys	0.10/0.071	
	Girls	<b>0.19</b> /0.001	Post- > Delayed
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.10/0.630	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.09/0.990	
<b>Difference between genders, Delayed-test</b>	Boys v Girls	0.06/0.282	

## Control Experiment

Factor: Awareness of outdoor environment

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases (117)	0.02/0.758	No statistically significant signals
	Boys (68)	0.06/0.523	
	Girls (49)	0.03/0.793	
<b>Difference between genders, Pre-test</b>	Boys v Girls (68 v 49)	0.15/0.096	
<b>Difference between genders, Post-test</b>	Boys v Girls (68 v 49)	0.16/0.085	

Factor: Perception of Competence

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	0.0720.271	No statistically significant signals
	Boys	0.06/0.502	
	Girls	0.10/0.347	
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.10/0.259	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.12/0.194	

Factor: Perception of Confidence

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	0.05/0.412	No statistically significant signals
	Boys	0.17/0.053	
	Girls	0.13/0.210	
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.04/0.629	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.08/0.367	

Factor: Awareness of Others

Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	0.00/0.957	No statistically significant signals
	Boys	0.07/0.387	
	Girls	0.13/0.217	
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.14/0.121	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.05/0.606	

Factor: Attitude towards Groups

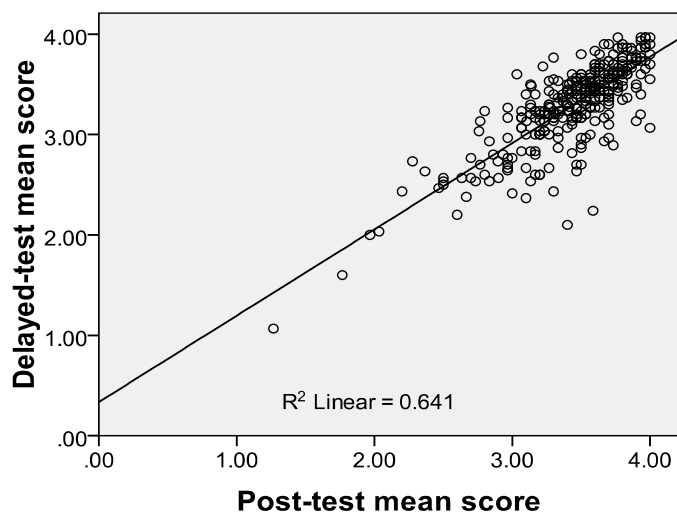
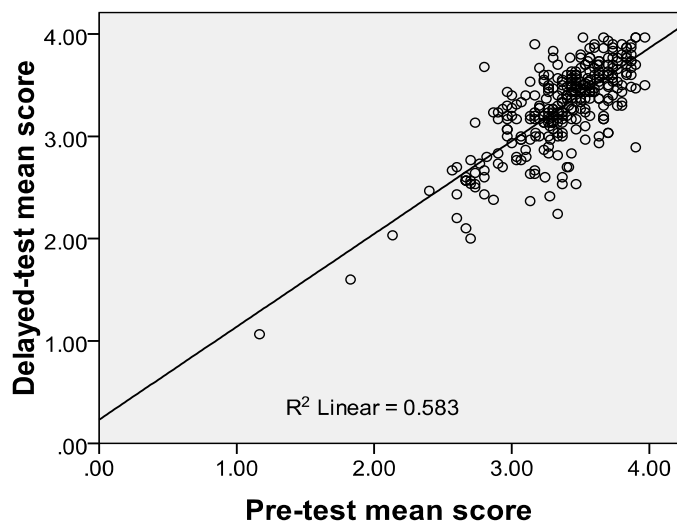
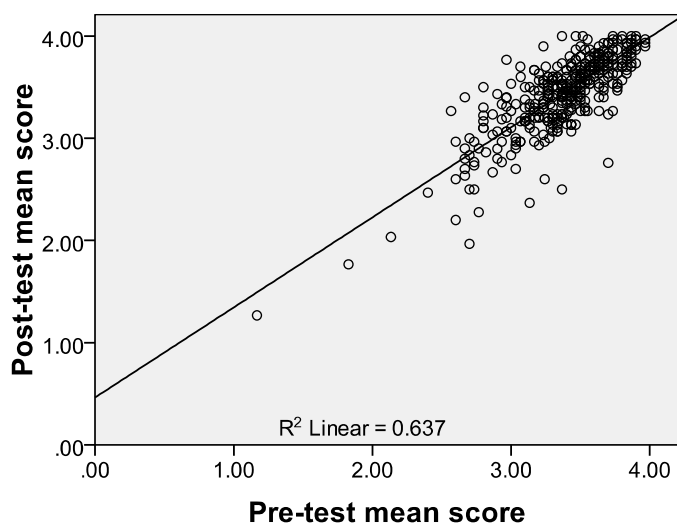
Test	Sample	r/p	Comments
<b>Difference in mean scores Pre-test – Post-test</b>	All cases	0.02/0.758	No statistically significant signals
	Boys	0.04/0.622	
	Girls	0.12/0.241	
<b>Difference between genders, Pre-test</b>	Boys v Girls	0.10/0.276	
<b>Difference between genders, Post-test</b>	Boys v Girls	0.02/0.811	

### **Appendix 3.4    Linear regression graphs for Pre-test vs Post-test vs Delayed-test mean scores for the Main Study and Control Experiment, for all cases and for boys and girls.**

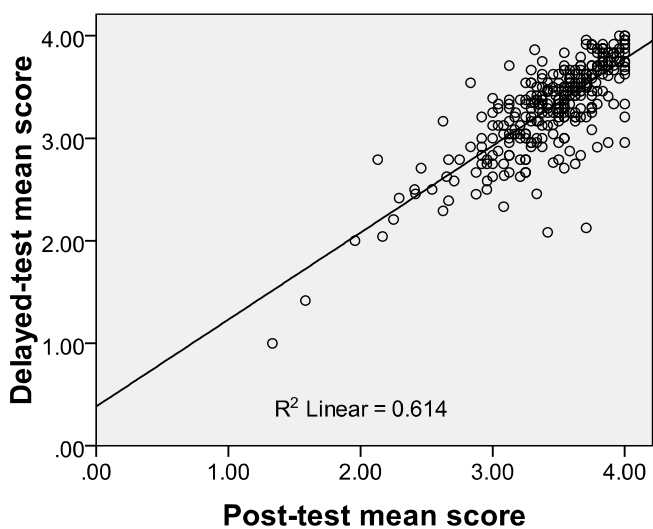
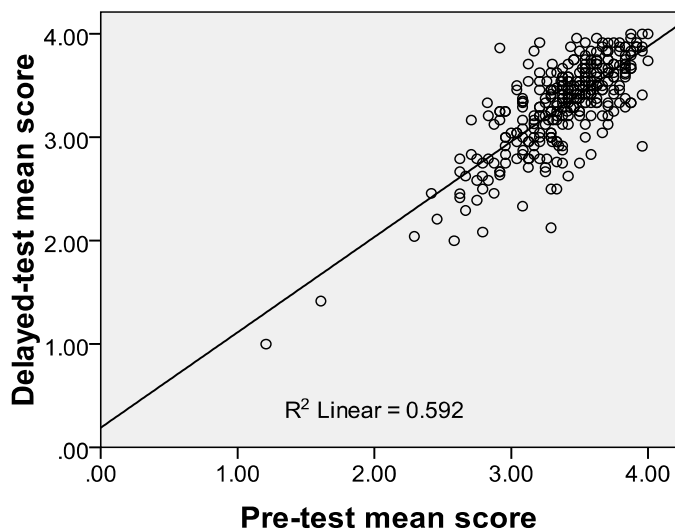
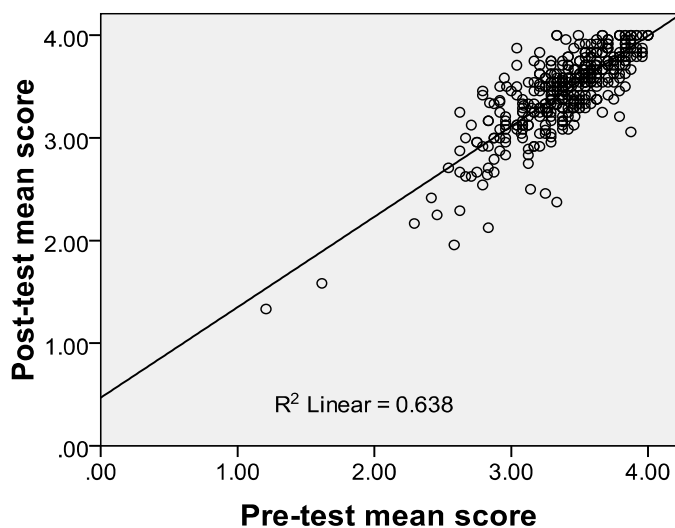
#### **Notes**

1. The graphs are grouped together for all items in the questionnaire, PSD items and items on “Awareness of the Outdoor Environment”.
2. The linear fit to the scatter plots is made by the least-squares, best-fit method.  $R^2$  is the Coefficient of Determination, a measure of how well the independent variable predicts the outcome in the dependent variable.

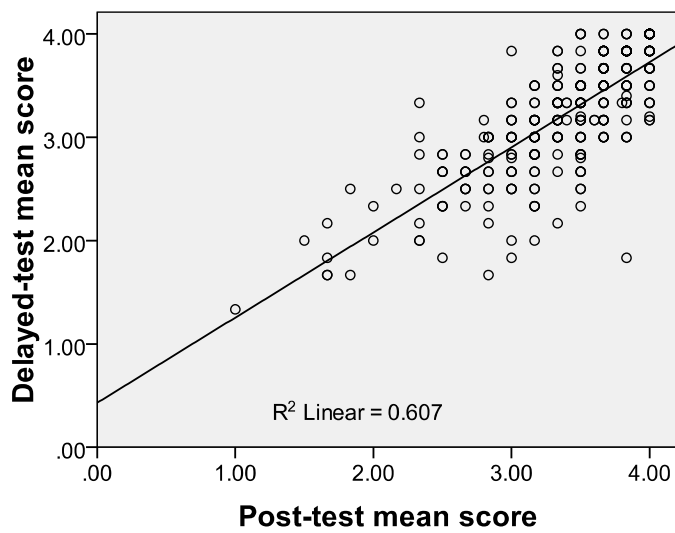
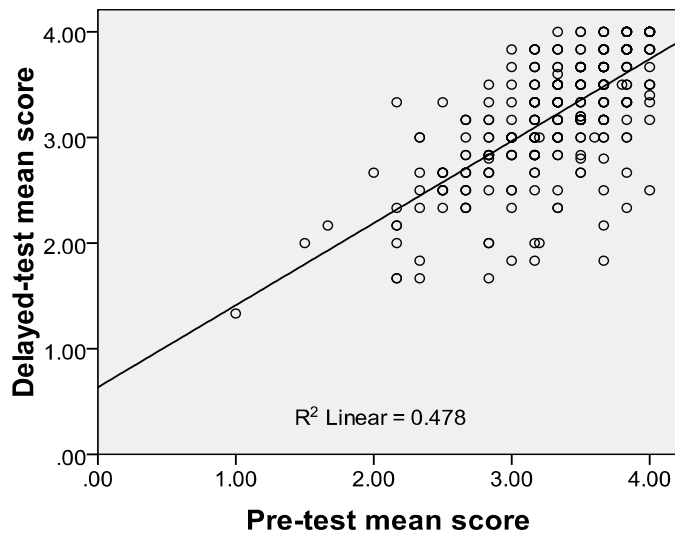
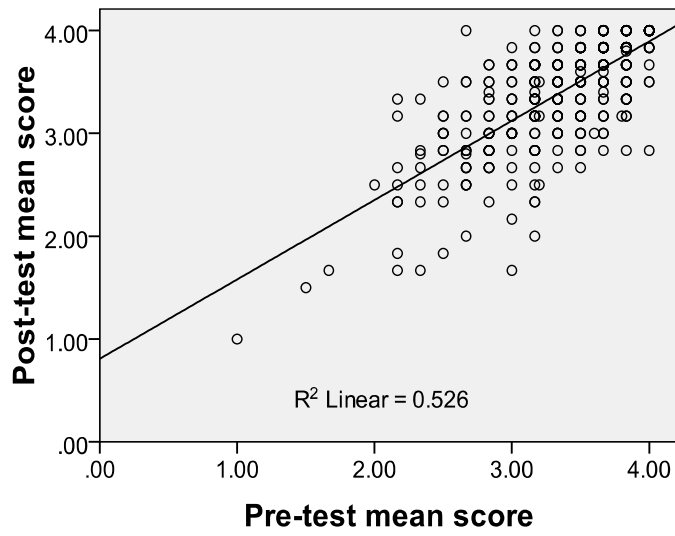
## Main Study, all cases, all items



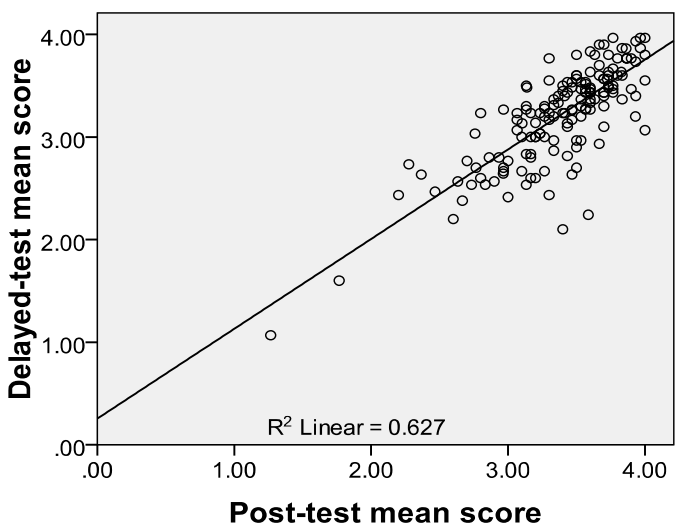
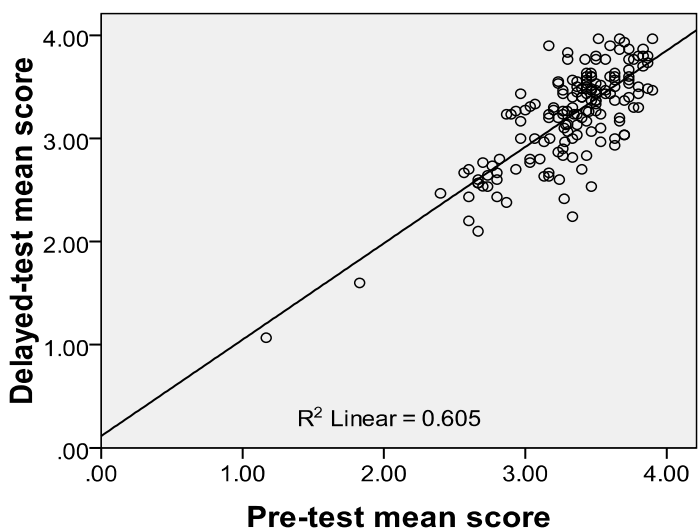
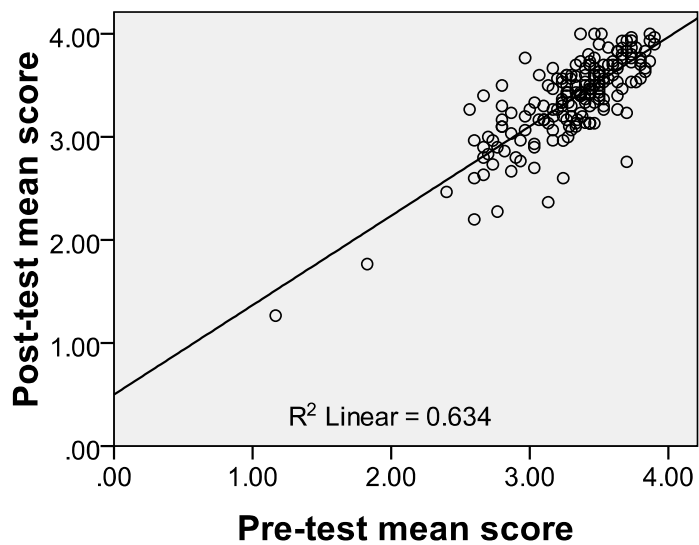
# Main Study, all cases, PSD items



**Main Study, all cases, “Awareness of the Outdoor Environment”**

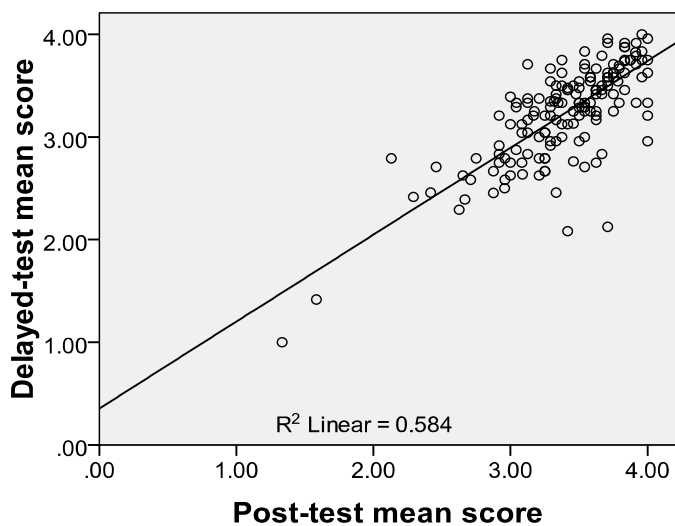
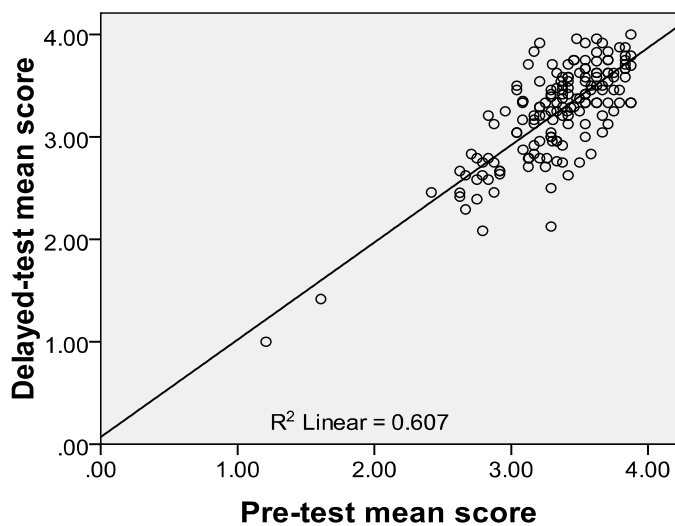
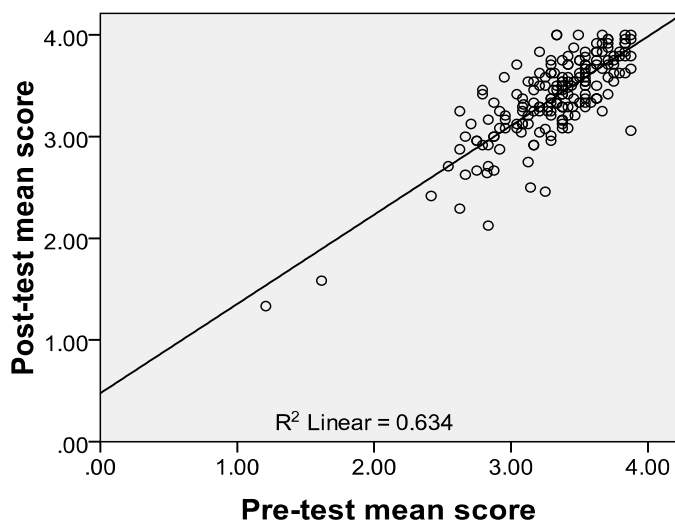


**Main Study, boys, all items**

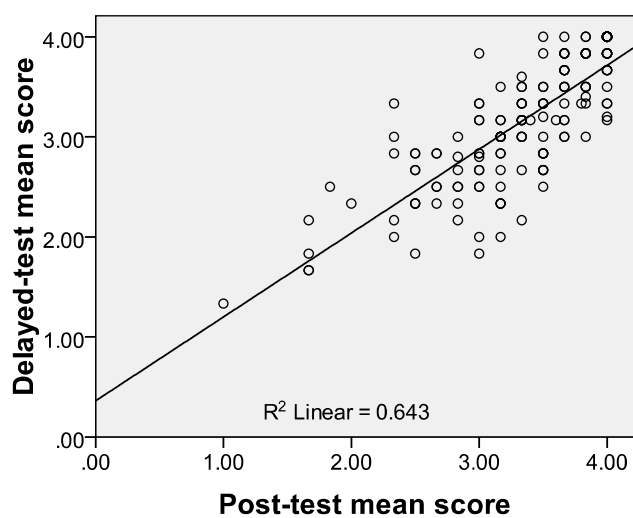
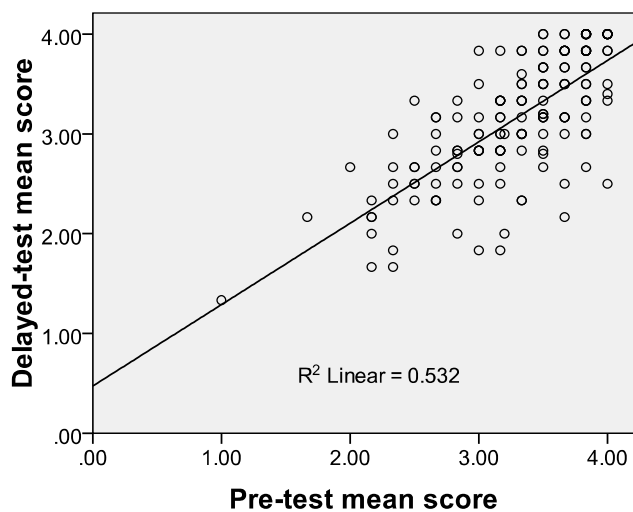
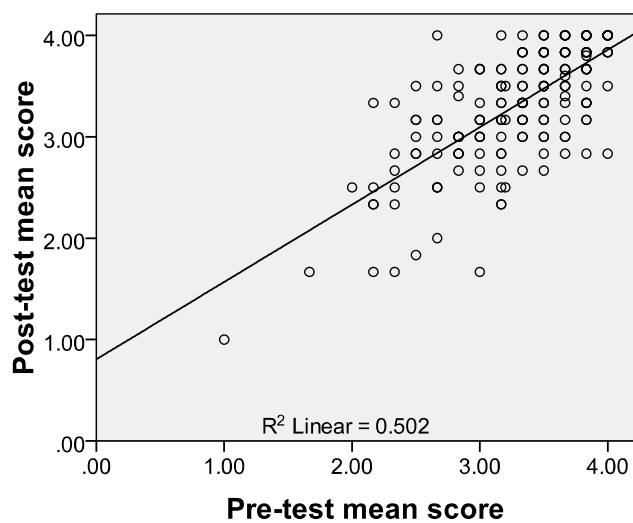




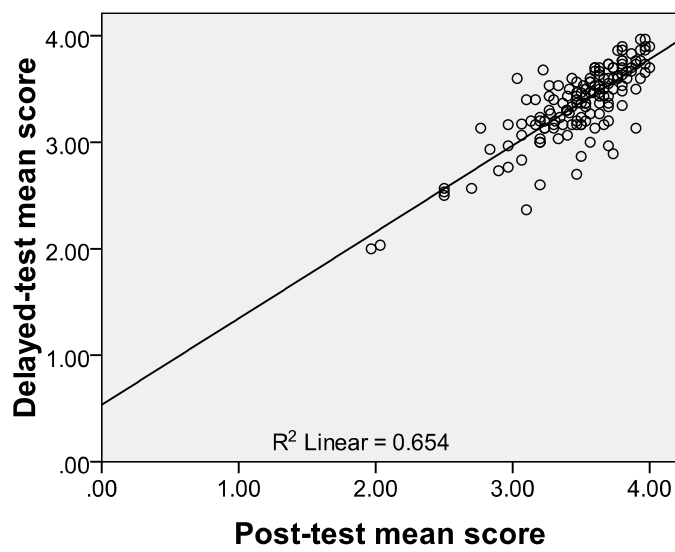
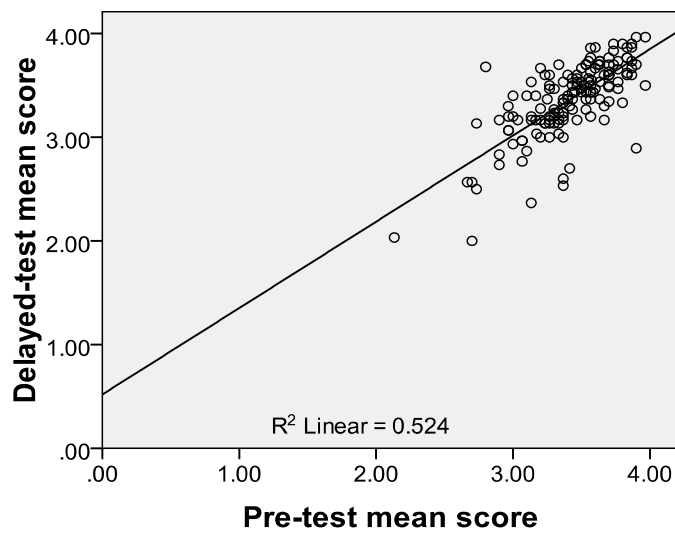
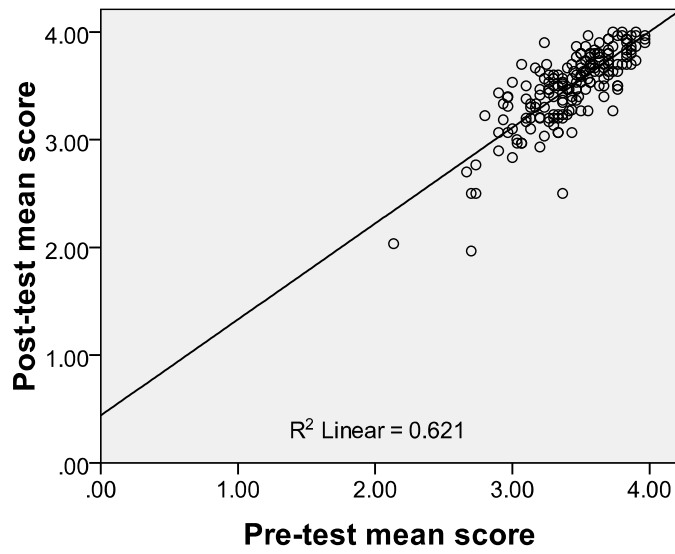
# Main Study, boys, PSD items



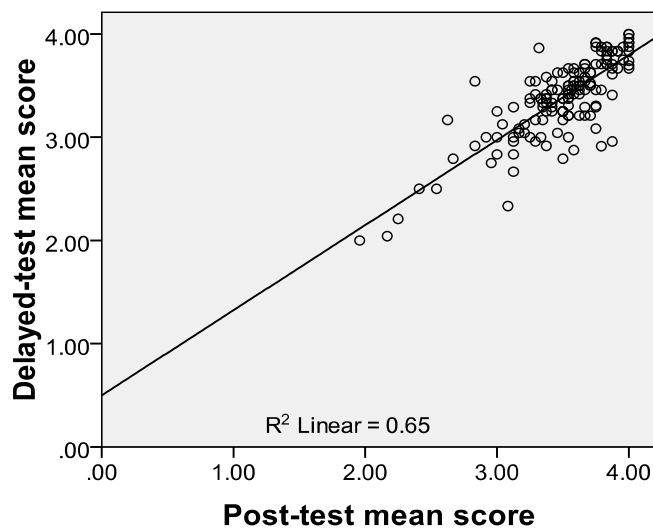
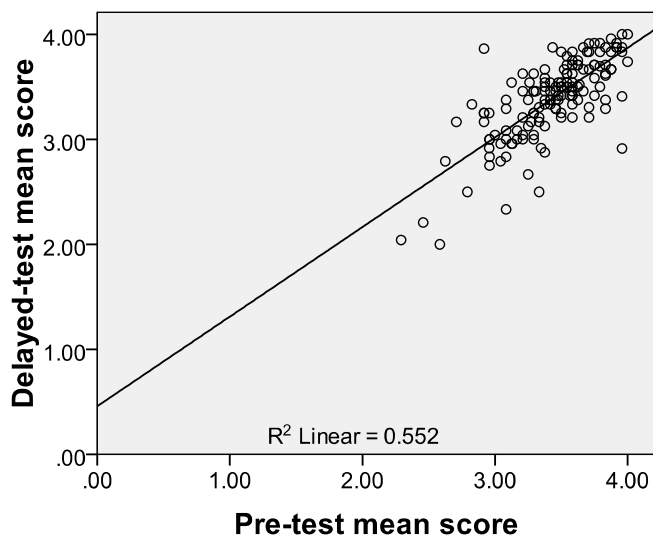
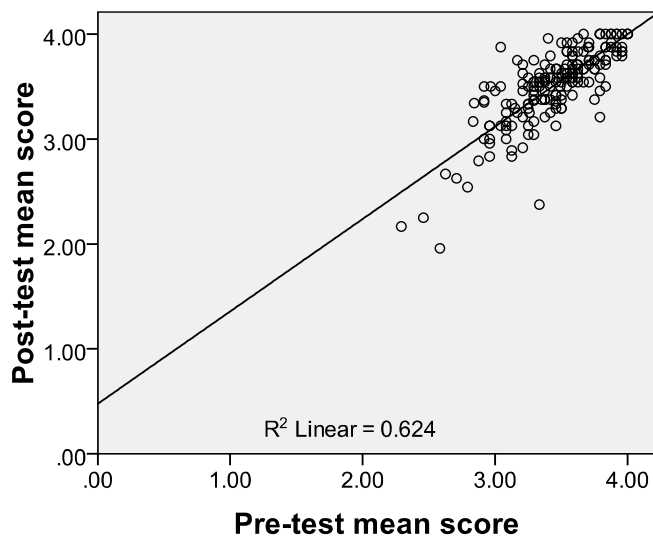
# Main Study, boys, “Awareness of the Outdoor Environment”



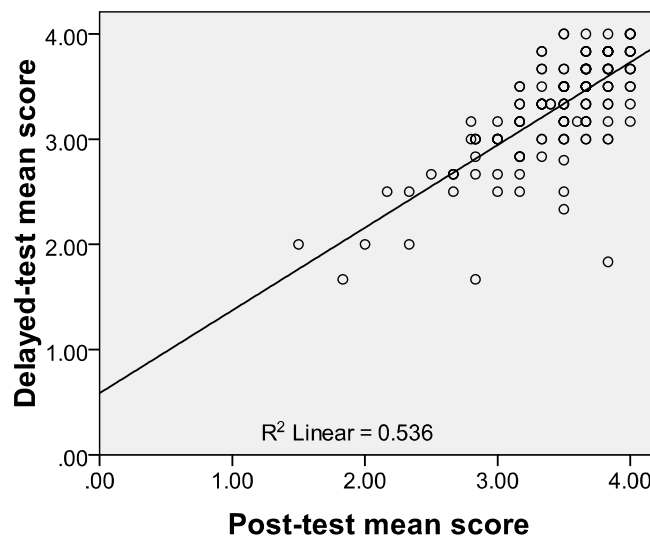
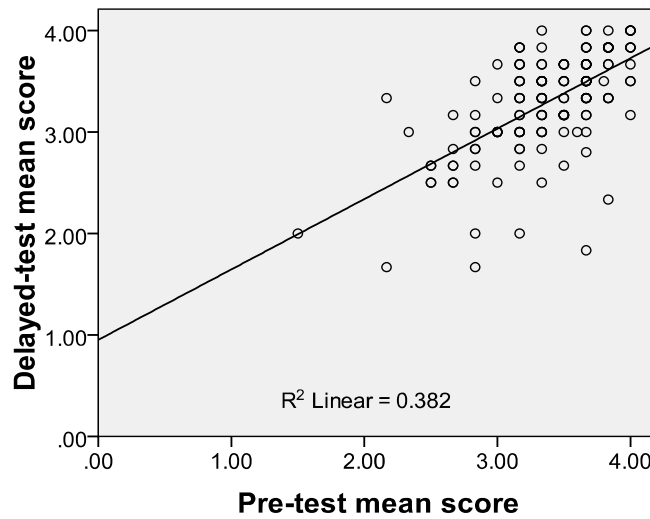
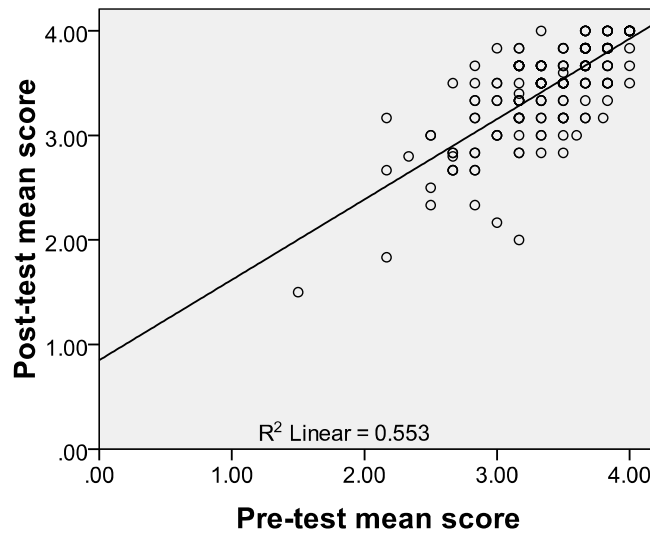
**Main Study, girls, all items**



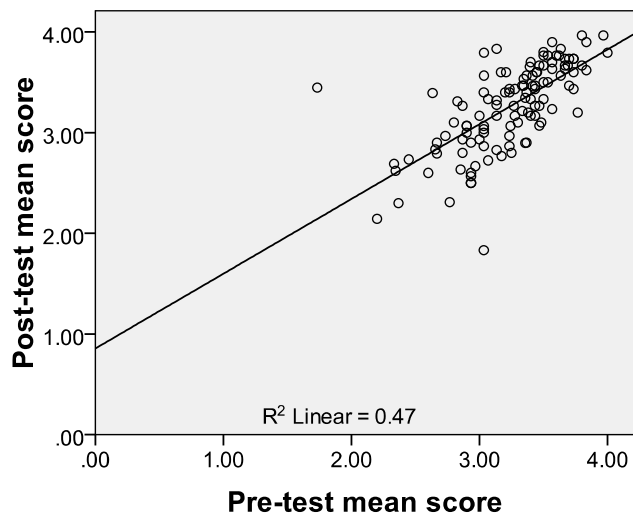
## Main Study, girls, PSD items



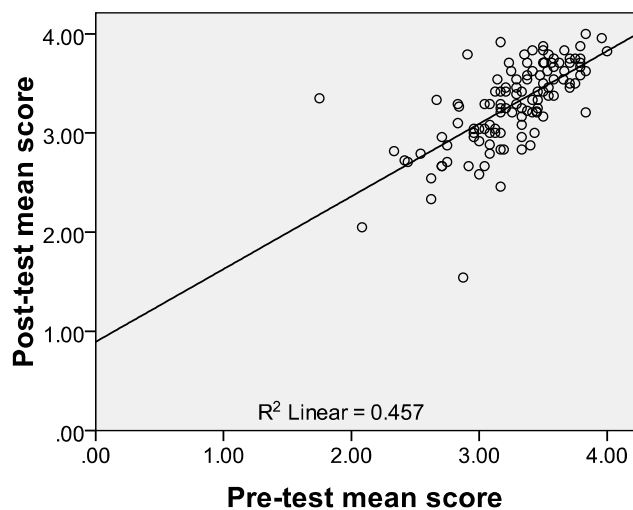
# Main Study, girls, “Awareness of the Outdoor Environment”



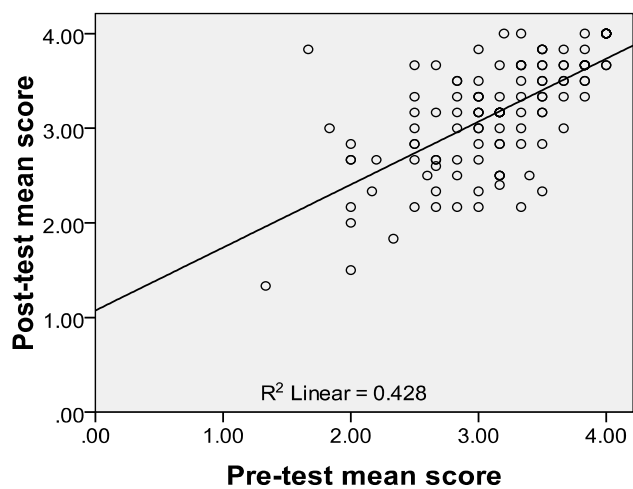
**Control Experiment, all cases, all items**



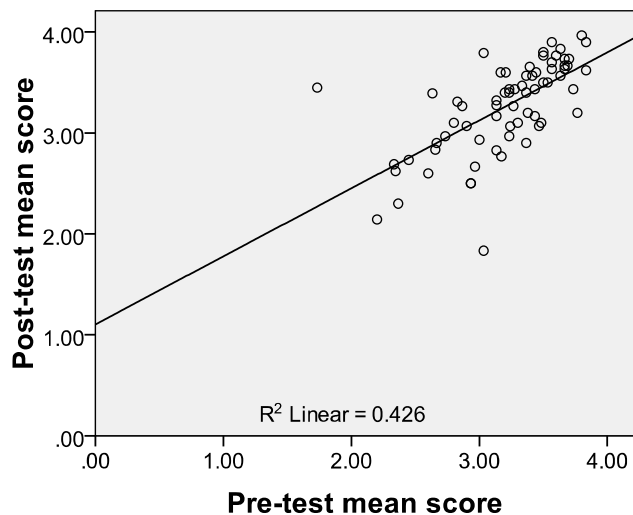
**Control Experiment, all cases, PSD items**



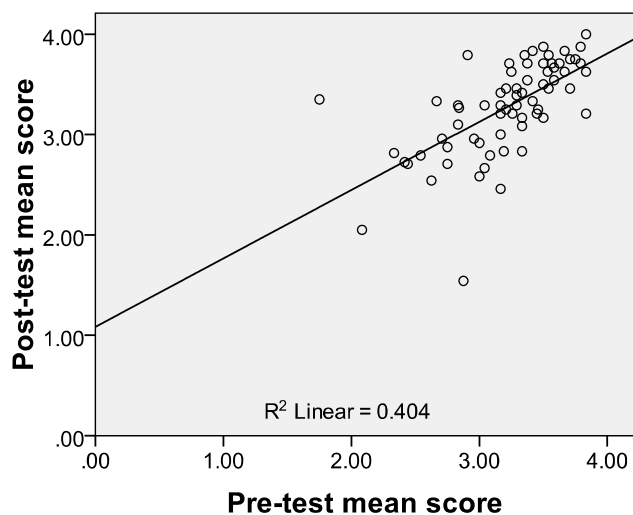
**Control Experiment, all cases, “Awareness of the Outdoor Environment”**



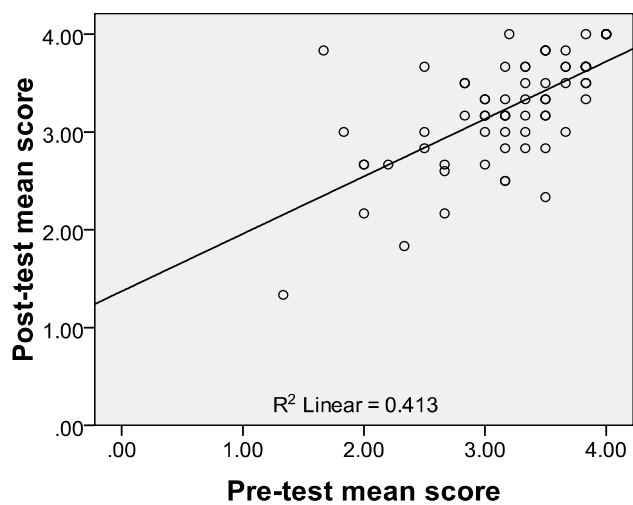
**Control Experiment, boys, all items**



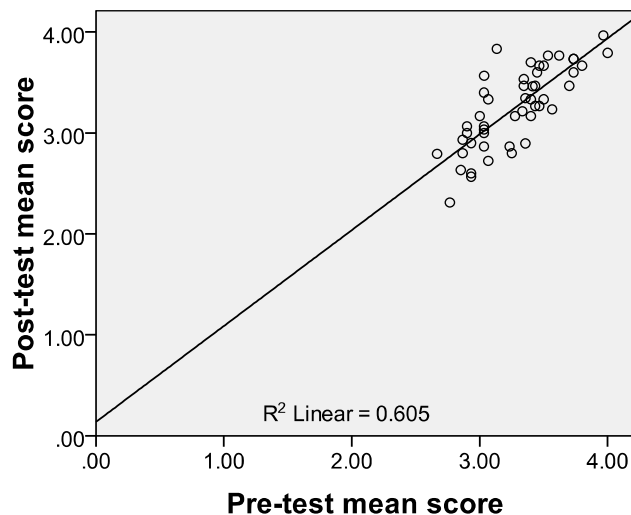
**Control Experiment, boys, PSD items**



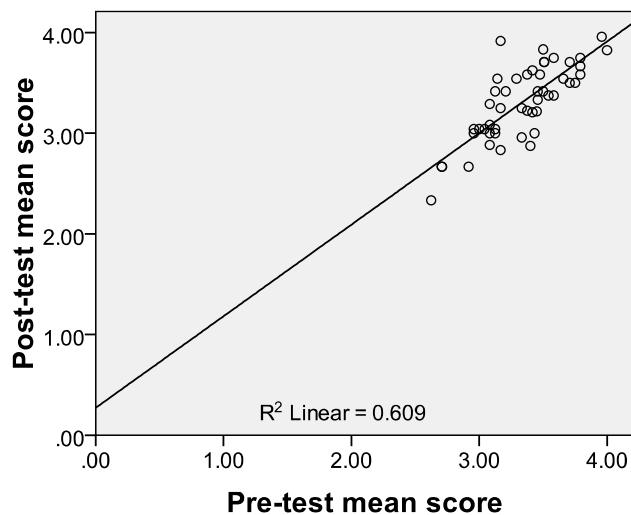
**Control Experiment, boys, "Awareness of the Outdoor Environment"**



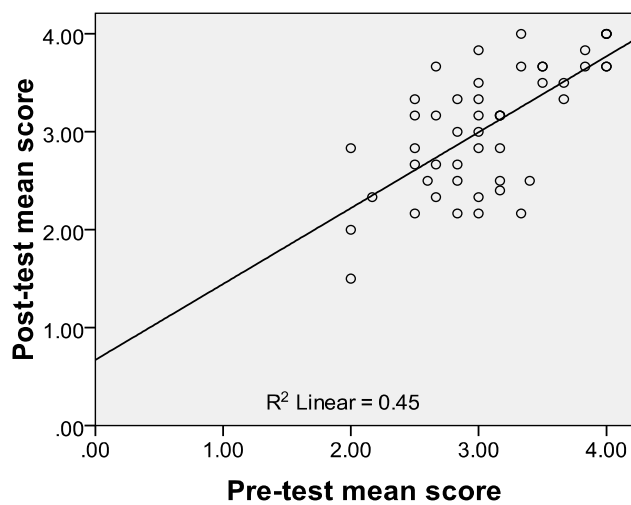
**Control Experiment, girls, all items**



**Control Experiment, girls, PSD items**



**Control Experiment, girls, “Awareness of the Outdoor Environment”**





## **Appendix 4.1      An understanding of children's learning process during outdoor adventure activities**

This Appendix underpins the suggestion in Chapter 4, Section 4.2.3 that theories of the learning process should be borne in mind when choosing the methodology for an empirical research project and for interpreting its outcomes. This seems self evident, but there are only a few indications that this has been done in published outdoor education research. This is despite a good foundation for such considerations being laid by Walsh and Golins (1976) and Hopkins and Putnam (1993). In a recent exception to this trend, Dillon (2005) argues that to understand and improve learning in outdoor education we must look to learning theories, concepts and perspectives. Perhaps it is because it is so obvious that recourse to theory should be made that it is rarely mentioned. To be honest, through lack of experience on my part it was not done in this project, and it was only when it was necessary to find reasons for the poor reliability of the new, tailored questionnaire and the large scatter in the data set, that theories of learning as they apply to 10-12 year old children emerged as relevant. What follows is a novice account of the theories and models of learning that underpin PSD from OAA.

To explain learning through outdoor adventure, extensive use has been made of Kolb's model of an experiential learning cycle (e.g. Hopkins & Putnam, 1993; McLoughlin, 1997; Higgins & Nicol, 2002). Kolb's model (Kolb, 1984) stems from several major and fundamental theories of learning, most notably experiential learning theory, and, at the level of a general concept, is widely appreciated and widely used in research. On the other hand, it should be noted that it has also attracted a good deal of criticism when analysed at a greater level of detail (Greenaway, 2010). Like so many models, there are less well understood specific cycles within the generic cycle. Nevertheless, it goes beyond the step of simply linking PSD outputs to PSD inputs, which is what has been presented so far in this thesis and in numerous other studies of the benefits of outdoor adventure education. Here, a short discussion is presented to justify the point that in order to understand why outdoor learning manifestly yields substantial benefits for school children it is important for research to take theories of the learning into account, go beyond simply linking outputs to inputs and seek to interpret empirical evidence in terms of the learning processes involved.

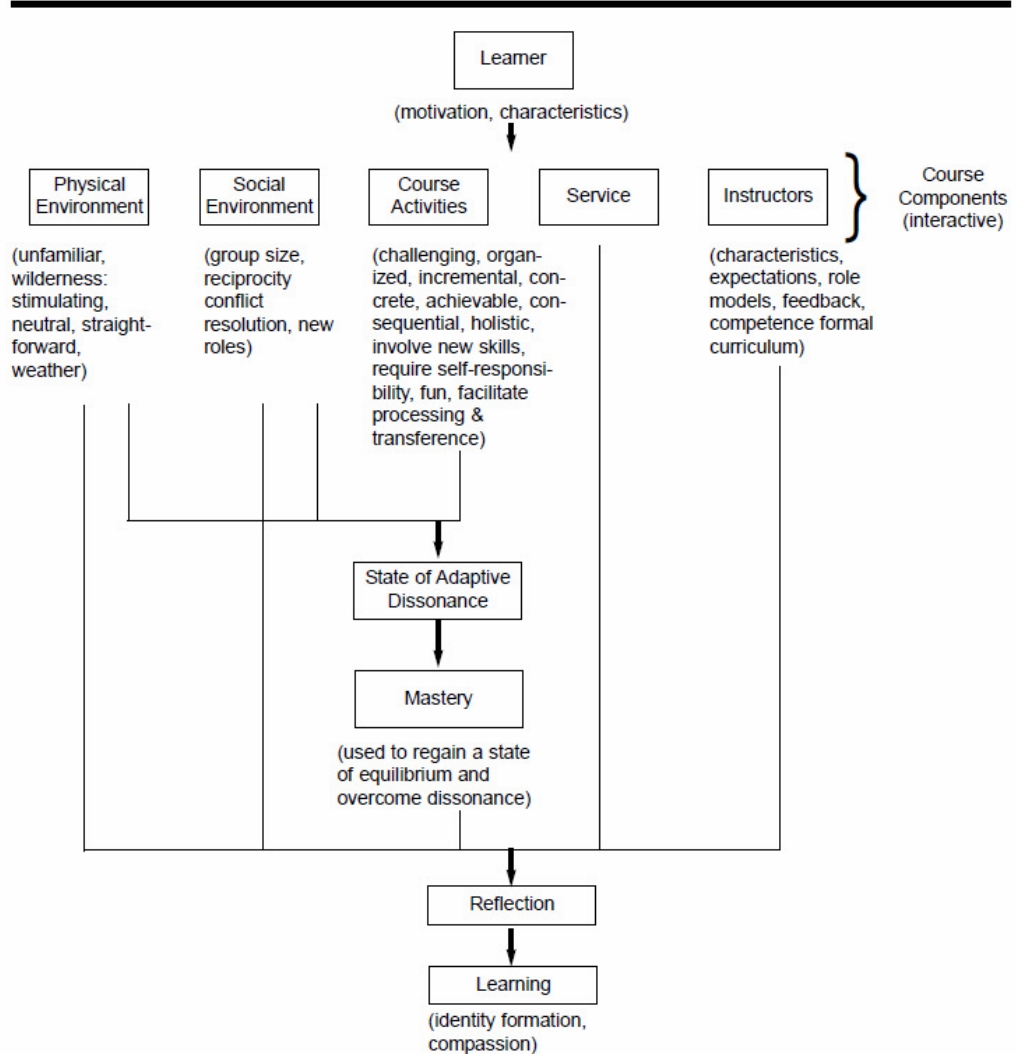
## From inputs to outputs

Barrett & Greenaway (1995), in their research review into the role and value of outdoor adventure in young people's personal and social development, wrote, "To understand the ways in which adventure contributes to young people's development, it is necessary to investigate the various *processes* from the individual perspectives of young people" (my italics) and, "No research has been found which systematically traces the various ways in which adventure affects individual young people's development". 15 years on, and from my reading of more recent literature, these statements still seem to reflect the situation in much of this area of research. It is the case that several studies have investigated further than simply measuring the benefit of OAA for PSD and have gone on to address the question of why outdoor adventure activities are so beneficial. However, the discussions commonly focus on the effectiveness of components of the experience, such as its duration, the contents of the course, the ethos behind the delivery, the background of the participant, follow-up to the experience and so on, rather than on the learning processes that convert these components into personal and social benefits. That is to say, the outputs are equated to the inputs, sometimes through statistical modelling, but without analysis of the conversion (learning) process.

Good examples of studies that have evaluated the components of the experience in this way are Neil & Richards (1998), McKenzie (2000) and Sibthorp (2003). Sibthorp actually points out, "it is critical that ... closer examination is afforded to the process behind adventure education" and then, "programming decisions remain largely an enigmatic process based on gut instinct, past experience, and borrowed or untested philosophical understanding or belief." He goes on to review briefly the Walsh & Golins (1976) adventure education model, which does consider learning processes at work during adventure activities, with the theory of adaptive dissonance (essentially cognitive dissonance, but emphasising the adaptive elements of it) at the heart of them. McKenzie (2003) takes a more critical look at the Walsh & Golins model. Although much of her discussion reverts to the effectiveness of the components of the experience rather than the learning process underpinned by learning theory, she does suggest that a model of effects or processes operating in parallel rather than the linear model of Walsh & Golins, would be more appropriate. Her Figure 2 is included here to illustrate what is probably the most advanced input-output (object) model of student learning from OAA that we have at the moment (Figure A4.1.1).

**Figure A4.1.1** A parallel object-oriented model of student learning during OAA.  
Figure 2 from McKenzie (2003).

**Figure 2. Building on Walsh and Golins' (1976) Outward Bound process: An alternative model of student learning.**



### From input-output to process - for the PSD of 10-12 year olds

It seems that what many researchers in outdoor education mean when they refer to a model of the learning process is, in reality, an object model, linking input objects to output objects, through statistical analysis where a quantitative approach is taken. Ewart's (1989) use of a causal model to relate causes and effects is an example of this. A process model should emphasize the individual's learning processes based on learning theories, which are well established in education and increasingly well understood through the research field of

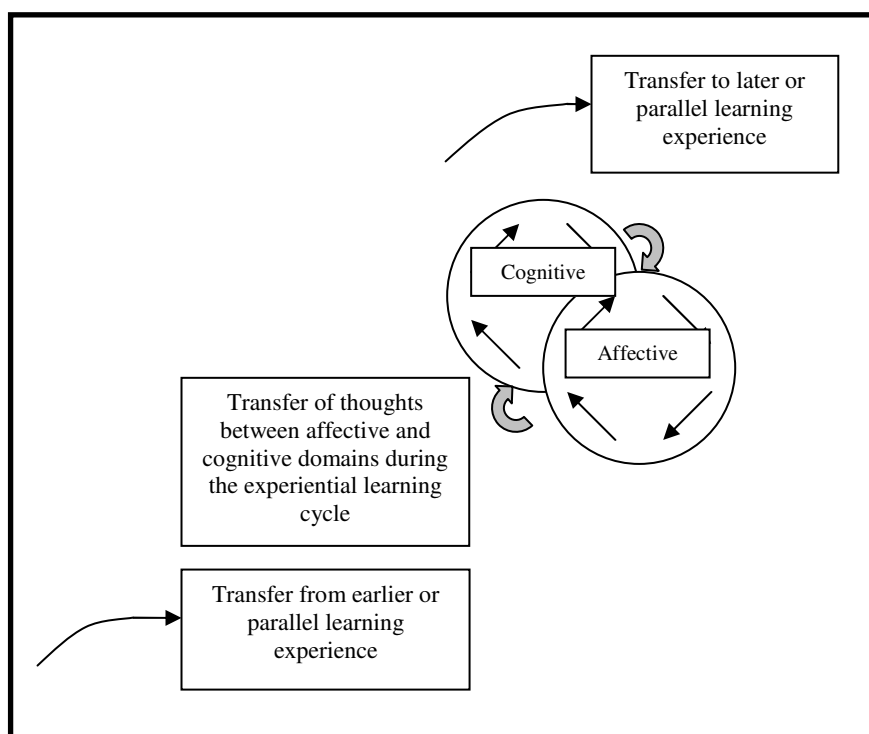
cognitive psychology, itself informed by connectionist models of learning and modern developments in neuroscience. It is acknowledged that programme components and the participant's traits and the way they interact are important for realising the benefits of outdoor education, but these interactions will only be fully effective in producing positive outcomes if they also take account of learning processes (Dillon *et al.*, 2005). Having said that, the building of object models, based as they are on the statistical analysis of observational data with little or no underlying theory, improves our understanding and informs the development of process models.

There are numerous theories of the learning process. From within these theories there are some that are particularly relevant to outdoor education, such as theories of experiential learning (e.g. Kolb, 1984), and from these some that will be particularly relevant to maximising the benefits of OAA for the PSD of children in the late primary years of their education, i.e. at 10-12 years old. For the purposes of applying theory to models of the learning process, the age of the learner is perhaps the most important variable once the area of learning and the learning environment are defined. This is because at one stage or another, nearly all explanation of the learning process refer back to Piaget's theory of cognitive development, which charts the development of the learner against age (Gross, 1992, Chap. 25 is a summary). Much subsequent research makes the point that the actual timing of development is strongly dependent on the individual learner and even on the domain in which learning is taking place (Bloom, 1956; Bloom *et al.*, 1964). Nevertheless, there is general agreement that at approximately 10-12 years old children are going through the transition from Piaget's concrete operational learning stage to the formal operational learning stage. At this critical transition for the children studied here, they are developing their learning processes by building on ones that manipulate objects or material things, as younger children do, to include those that manipulate ideas, propositions or concepts, as adolescents and adults do. Included in the latter is the ability to think constructively about situations they have not actually experienced before, which could be particularly important during a first outdoor adventure activity experience. Also, and crucially important for PSD, at this same transition, in Piaget's opinion, children develop the ability to recognise and consider their own and others' behaviour and the potential differences between these.

However, the learning of personal and social skills takes place in the affective domain rather than the cognitive domain of Piaget, and the debate about the relationship between cognitive and affective thought processes becomes important. In reviewing the concepts of cognitive

appraisal and affective primacy both Gross (1992) and Eysenck & Keane (1998) conclude that cognitive appraisal to some extent or other precedes an affective response to a stimulus. This situation is exemplified by Walsh and Golins' (1976) use of adaptive dissonance theory, since dissonance must be recognised by the individual before "anxiety" sets in and s/he adapts to "reachieve harmony" (Walsh and Golins' terminology), although whether this is done at a level of conscious awareness is still a matter for debate (Steinberg, 2005). Bloom *et al.* (1964) felt that both cognitive and affective domains comprise a sequence of thought processes that are "tightly intertwined". If this is so, then the transition from concrete to formal operational learning in Piaget's cognitive development theory impacts on what it is possible to learn in the affective domain. Moreover, the essentially cyclical nature of Piaget's cognitive development imposes on these thought processes a cyclical nature, and through this we arrive at Kolb's model of experiential learning. We can extend this model by adding the concept of transfer of learning (Haskell, 2001), which is also implied in Piaget's theory of cognitive development. Figure A4.1.2 summarises the essential points of this kind of model. At the transition from concrete to formal operations, 10-12 year old children are becoming capable of exploiting this cycle, or process, especially having been introduced to the necessary PSD concepts in the school curriculum.

**Figure A4.1.2** Simple graphical representation of a learning cycle incorporating "intertwined" (shaded arrows) cognitive and affective experiential learning embedded in a linear or parallel set of learning experiences.



## Why learn so quickly?

Even if the learning of personal and social skills during a residential week of OAA can be understood in terms of current theories of the learning process, an intriguing aspect of the phenomenon is the speed at which it appears to happen. Such skills are what we might call “life skills” and we would expect them to be learnt over months if not years, rather than over the few days of a residential week. If they are learnt so quickly, then with OAA and PSD, are we witnessing an example of “accelerated learning”?

Although unexpected in social skills, there is some evidence that learning social concepts can be quite rapid. McNaughton (1966), in a study on the implications of Piaget’s research for the teaching of social studies in primary school, points out that a rich experiential environment accelerates the process. Quite remarkably, a group of 10-12 year old school children in a story-telling session increased the ratio of abstract (formal) concepts to concrete concepts in their answers to questions from the teacher from 1:4.0 to 1:1.3 within the scope of one classroom session. From reference back to Figure A4.1.1 it is clear that residential OAA has the potential to provide such a rich experiential learning environment with several parallel threads of both social and non-social kinds. The presence of social as well as non-social threads potentially brings into play the Developmental Theory of Vygotsky (Gross, 1992), which sees the increasing internalisation of an individual’s thinking and reasoning as a fundamentally social process. Therefore, it would seem that OAA residencies are an ideal environment for accelerated learning, richer in experiences than the classroom and even richer than learning outside the classroom on a non-residential basis. There is also some evidence that certain physiological changes promote accelerated learning. Shephard (1997), in reviewing the impact of regular physical exercise for primary school children on academic performance, concludes that the resultant increased cerebral blood flow and greater arousal are amongst factors that help “learning ... to proceed more rapidly per unit of ... time”. Although the time scales over which these effects operate might vary, again it seems that OAA has the potential to provide these kinds of stimuli.

In outdoor education accelerated learning has, indeed, been widely observed, but largely in the form of a “fieldwork effect” occurring in the cognitive domain (Bogner, 1998; Nundy, 1999). Nundy appeals to memorable incidents occurring during residential geography fieldwork as triggers for the improved recall of knowledge. An interesting question (although a slight digression from the main thread of this discussion) is whether residential

fieldwork for cognitive learning delivers a simultaneous fieldwork effect with associated accelerated learning for personal and social skills, in much the same way as OAA appears to do. Unfortunately, in comparison with studies of the impact of OAA on PSD, there are few studies on which to base an answer to this question. Possible examples for pre-adolescent children can be found in Mygind (2009) with a non-residential experience at a Copenhagen Folkeskole, in Bogner (1998) with a five-day ecology programme in the Bavarian National Forest Park, and in Dillon *et al.* (2005) with a residential ecology course in the New Forest, but in all these cases it is not clear whether the affective learning was in academic or social areas. By far the most interesting study is Nundy's geography fieldwork with late primary children. While pupils showed improvement in the affective domain for their relationship with their academic work there was virtually no improvement in the affective domain for their social skills. A self-concept questionnaire was used to score the effects and a comparison made with a control group. This table gives examples of the questions used to measure the benefits of the geography fieldwork:

<b>Academic cognitive</b>	<b>Academic affective</b>	<b>Non-academic affective</b>
<i>I'm good at geography</i>	<i>I like geography</i>	<i>I am an attractive person</i>
<i>I get good grades at geography</i>	<i>I enjoy doing work for geography</i>	<i>I get along well with my parents</i>
		<i>I like to run and play hard</i>
		<i>I make friends easily</i>

Nundy's t-test results for the difference in mean scores between the experimental and control groups of pupils in the affective part of the questionnaire (two right-hand columns above) were:

<b>Self-concept questions</b>	<b>t-test result</b>	<b>Effect size*</b>
All questions	1.88	0.20
Academic affective questions	1.95	0.21
Non-academic affective questions	0.75	0.08

\*Effect size was not reported by Nundy but has been calculated here from the equation,  $r = \sqrt{t^2 / (t^2 + df)}$ , where  $t$  is from "t-test result" and  $df$  is the sample size (Field, 2009, p.340).

Nundy does not comment on the low t-test result of 0.75 ( $r=0.08$ ) for non-academic (social) affective learning, but it is clear that there is a strong difference from the academic scores. We can conclude from this digression that any evidence for residential field courses in the late primary years yielding a fieldwork effect for PSD, with associated accelerated learning, is mixed at best. Rickinson *et al.* (2004) reviewed this area and much the same conclusion can be reached from looking at the publications they used.

## Conclusions

Residential outdoor adventure activities offer something that field courses and other forms of outdoor education do not offer in terms of the rich experiential environment on which the rapid development of personal and social skills seems to depend. The object (input-output) model of McKenzie (Figure A4.1.1) describes the structure of this rich environment (or, at least, the Outward Bound version of it). It is an environment that few, if any, of the 10-12 year old pupils studied here will have encountered before. The exceptionally wide range of emotions experienced under such circumstances is unique. Equally important is the role of the skilled facilitator to exploit this environment and create opportunities for the learning process to function. His/her job facilitates the integration of the concepts illustrated in Figures A4.1.1 and A4.1.2, so that experiential learning is transferred between parallel elements of the residential experience. By way of background to this experience, children of 10-12 years old are undergoing profound changes in their cognitive learning skills as they progress from concrete-learner to formal learner and the links this has with learning in the affective domain. When the rich experiential learning environment of the residential week is embedded in the concurrent transition to higher learning skills the result is dramatic.